DOCUMENT RESUME

ED 169 189 UD 019 292

AUTHOR	Carrillo, Federico Martinez, Ed.
TITLE	Approaches to Assessing Bilingual/Bicultural Programs.
INSTITUTION	New Mexico Univ., Albuquerque. American Indian Bilingual Education Center.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
BUREAU NO	403GH80006
PUB DATE	Jan 79
GRANT	G007804907
NOTE	98p.
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	American Indians; Biculturalism; *Bilingual Education; *Compensatory Education Programs; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Evaluators; Intercultural Programs; Minority Groups; *Program Evaluation
IDENTIFIERS	*Elementary Secondary Education Act Title VII

ABSTRACT

The papers in this monograph reflect different approaches used in the assessment and evaluation of bilingual education programs. Doctors Harry Berendzen and Katherine Bemis provide some practical considerations for implementation of evaluation. They list responsibilities of an evaluator and offer an evaluation management plan which takes into consideration needs assessment, performance objectives, measurement of project impact, material development, staff development, and community involvement. Dr. Michael Massarotti offers a systematic procedure for collecting information based on pre-established objectives and subsequent analysis, reporting, and use of the information for the purpose of making decisions concerning program development. Dr. Charles Leyba provides an overview and design of Title VII program evaluation. He also discusses standardized testing and analysis of scores. Finally, the topics of Dr. Robert Reeback's paper include a critique of Title VII regulations concerning evaluation, some examples of misguided or inadequate evaluation, a review of some persuasive evaluation results, and suggestions for reconstituting evaluation in the context of Title VII and adjusting to it in the meantime. (Author/EB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED169189

APPROACHES TO ASSESSING BILINGUAL/BICULTURAL PROGRAMS

edited by

Federico Martinez Carrillo
Coordinator of Assessment
and Evaluation
January 1979

American Indian Bilingual
Education Center
College of Education
University of New Mexico
Albuquerque, New Mexico 87131

UD019292

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

F. Carillo
AIBEC

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE PROJECT REPORTED HEREIN WAS PERFORMED PURSUANT TO A GRANT FROM THE U. S. OFFICE OF EDUCATION, DEPARTMENT OF HEALTH, EDUCATION AND WELFARE. HOWEVER, THE OPINIONS EXPRESSED HEREIN DO NOT NECESSARILY REFLECT THE POSITIONS OR POLICY OF THE U. S. OFFICE OF EDUCATION AND NO OFFICIAL ENDORSEMENT BY THE U. S. OFFICE OF EDUCATION SHOULD BE INFERRED.

TITLE VII PROJECT
GRANT #G007804907 - PROJECT #403 GH80006

THIS PUBLICATION WAS DEVELOPED AND PRINTED WITH FUNDS PROVIDED BY THE BILINGUAL EDUCATION ACT, TITLE VII OF THE ELEMENTARY AND SECONDARY EDUCATION ACT OF 1965, AS AMENDED BY PUBLIC LAW 93.380. THEREFORE IT IS IN THE PUBLIC DOMAIN AND MAY BE REPRODUCED FOR LOCAL USE.

TABLE OF CONTENTS

List of Contributors iv

Preface v
Dr. Rudy Cordova

Acknowledgement viii
Dr. Federico Martinez Carrillo

1. Practical Consideration For Evaluating American Indian Bilingual Education Programs 1
Dr. Harry B. Berendzen
and
Dr. Katherine A. Bemis

2. Por Los Ninos 25
Dr. Michael C. Massarotti

3. Evaluation of Title VII Programs Overview and Design 44
Dr. Charles F. Leyba

4. Evaluating Bilingual Education Programs while remaining coherent 63
Dr. Robert T. Reeback

LIST OF CONTRIBUTORS

Dr. Harry B. Berendzen
and
Dr. Katherine A. Bemis

Educational Research Associates
Box 14896
Albuquerque, New Mexico 87131

Dr. Michael C. Massarotti

Executive Director,
Program Development and Research
4476 West 68th Avenue
Westminister, Colorado 80030

Dr. Charles F. Leyba

Director, Dissemination and Assessment
Center for Bilingual Education
California State University
5151 State University Drive
Los Angeles, California 90032

Dr. Robert T. Reeback

Program Evaluator
3029 Delano Place N.E.
Albuquerque, New Mexico 87106

PREFACE

Dr. Rudy Cordova
Director, Division of Program Development
Division of Bilingual Education
U. S. Office of Bilingual Education
400 Maryland Avenue S.W.
Reporters Building Room 421
Washington, D.C. 20202

Of all the reasons for evaluating bilingual education programs, the weakest is because the funding agency requires it as part of the grant. Bilingual education needs to be evaluated for the same reasons that other educational programs need to be evaluated -- to improve on what we are doing and to find better ways of doing a more effective job of educating children. Bilingual education, as an educational strategy, needs an on-going assessment of its effectiveness to determine the best techniques and approaches of instructing children in a language other than English. This is the part of the program that has been unsettling to some people in the education community. It is not so much that another language besides English has never been used to instruct children in this country, but rather that through Title VII of the ESEA larger numbers of the school population are being instructed in a language other than English, that is placing bilingual education on the front burner of the stove of critical analysis. It may be that this critical analysis of bilingual education will set a precedent for assessing other instructional modes that we in the education community have taken for granted simply because that is what schools have done for years. Regardless of the rationale behind the evaluation argument, our

approach as educators has to be one of using the best means available to provide the best answers possible to the question, "Why bilingual education?"

Education has always defied evaluation. The spectrum of human behavior is such that what may work for one individual will not work for another; what turns one child on to learning may turn another off; what may be relevant for one teacher may not be relevant to another teacher and sometimes what one person views as success another views as failure. Given this disparity of ideologies, one element that must always be constant is achievement. But not achievement based on one's own thinking, rather achievement based on measureable learning objectives of those for whom the bilingual program is intended. Bilingual/Bicultural education cannot be successful if children still fail to learn. So it must be that assessment and evaluation must take on several approaches. As one aspect of evaluation judges the impact of the administrator, another judges the impact of methodologies; one judges the effect of a learning environment, another judges the effect of attitudes. While approaches to assessment may be different, the ultimate goal of all assessment must be to relate all the parts to the over-all goal which is learning and this, translated, means the success of children.

The papers in this monograph are reflective of the different approaches used in evaluating a program. They present the reader with ideas on how different aspects of a program may be looked at. It is not intended to be an exhaustive report on assessment, but rather a compendium of ideas that may lead a person to more fully appreciate the magnitude of a field that

is sometimes taken lightly. Assessment and evaluation are not goals in themselves, but rather represent a tool which can help us determine how close we are coming to realizing those goals.

ACKNOWLEDGMENT

As a staff member of the American Indian Bilingual Education Center and editor of this monograph, I would like to express my appreciation to the many people who assisted in the original planning and organization of the monograph. In particular, I wish to thank my colleagues, Joe Gonzales and Margaret Wilcox, for their assistance, encouragement and support. The cooperation and thoughtfulness of the individual contributors greatly simplified the task of organizing and coordinating the articles for publication, and I wish to express my appreciation for their help. I am particularly grateful for the pleasant and meticulous manner in which Jean Minor and Nora Garviso worked on the preparation and typing of the final manuscript. In addition, I would like to thank Louis Baca for developing the art work and design for the cover. Finally, I wish to thank Dr. Rudy Cordova who, as always, provided enthusiastic support and assistance throughout the project.

PRACTICAL CONSIDERATIONS FOR EVALUATING AMERICAN INDIAN BILINGUAL EDUCATION PROGRAMS

Dr. Harry B. Berendzen

and

Dr. Katherine A. Bemis

Educational Research Associates
Box 14896
Albuquerque, New Mexico 87131

"I'm happy about the fact that our children have the advantage of bilingual education. When I use my own language, I say what I want to say, but when I use English, I am not always sure that I say what I want to." A PAC member of the Kerwo Title VII Bilingual Program (Papago/English) on the Papago Indian Reservation in Southwestern Arizona, expressed her feeling about the importance of language and thought. This same statement is applicable to children entering school speaking a language other than English and is probably multiplied in intensity and confusion.

Rather than addressing evaluation of Title VII Programs from a theoretical viewpoint, we will attempt to provide some practical considerations for implementation of evaluation. We specifically wish to relate these ideas to projects which are in a developmental process, and therefore, require both formative and summative systems of evaluation. By formative evaluation we mean stating project goals and objectives in a measurable way and providing regular feedback on their attainment so that the project may make modifications as it continues to develop. Summative evaluation refers to the sum total of data collected on student achievement, on

the attainment of objectives and other goals set up by the project. Data is compiled to determine how much impact the project has on student achievement over a period of time.

We consider the role of the evaluator to be that of a team member who works with the project staff to define needs, set up realistic goals and measurable objectives, and determine criteria to appropriately assess overall project effectiveness.

The role of the evaluator includes the following responsibilities:

1) Provide accountability in the use of federal funds for Title VII Bilingual/Bicultural programs.

2) Provide necessary information for program decision-makers to enable program modification to take place on an ongoing basis.

3) Provide information for consumers of the evaluation, such as parents, teachers, administrators, and other community members.

4) Assure that essential documentation is being collected for purposes of program continuity.

5) Provide a system to measure program impact on the education of children.

Special consideration should be given to the following areas as Program Directors begin to implement a total evaluation: an evaluation management plan, consisting of needs assessment, performance objectives, and measuring program impact; program components including instruction, staff development and acquitition, and community involvement.

The last section of this paper is a checklist which may be used by Program Directors to summarize optimal evaluation activities. Each component will be addressed in this paper with special consideration for the unique needs of a fully functioning Bilingual/Bicultural Title VII Program.

These components should interrelate and contribute to a program implementation which is compatible with a total program philosophy.

Most Native American Bilingual/Bicultural projects possess unique characteristics which require individual consideration. Most are in developmental stages of writing new orthographies for greater simplicity and transfer to English. Some languages are being written for the first time. Paraprofessionals are being trained to become responsible to the unique nature of their own language and culture. Parents are becoming more involved in the education of their children because they better understand the importance of education. Communities and schools are working as teams, mutually supporting each other's objectives.

EVALUATION MANAGEMENT PLAN

Needs Assessment

Prior to program planning, a needs assessment should be completed. The needs assessment offers documentation regarding the local education agency's unique characteristics and the disparity between these characteristics and the services offered by the schools meeting these needs. A need is defined as the difference between the current and desired situations. If a project can define, explain, and document its current situation and can also state where it wants to be, needs can be defined.

Program goals can then be established to meet the needs. The description of the current situation includes the objectives, cost, and assessments of the program as it exists while the desired program should take into consideration pupil needs, community, staff, and federal desires, expectations, and cost.

There are several methods which can be used to conduct a needs assessment. Some of these are as follows:

1. Reports. Interested groups may be asked to provide information regarding a given situation.

2. Tests. Results of standardized or criterion-referenced tests may indicate needs in an academic area.

3. Questionnaires. Questionnaires may be distributed to community members.

4. Census. A community census may be taken wherein a given target population is interviewed.

5. Goal Ranking. Interested persons are given program goals and/or objectives and asked to rank them according to need.

Performance Objectives

The needs derived form the basis for the program proposal in that program objectives are derived from the identified needs.

The importance of program objectives which are sometimes referred to as "behavioral objectives" was most significantly emphasized by Mager, 1962 in

his book, Preparing Instructional Objectives. Mager stressed the need to clearly delineate student or learner behavior, which for the evaluator means defining objectives in measurable terms. For example, a performance objective might state that "the staff will produce four newsletters during the project year to update the community on program activities" rather than simply, "the staff will keep the community informed."

The criterion of four newsletters is defined and can be measured against actual performance.

It is the Program Director's role to ascertain that a needs assessment has been conducted and that these needs are addressed through measurable performance objectives.

In specifying program objectives, the Program Director with the assistance of the administration, staff members, community members, and support staff should be cognizant of the way these objectives relate to the needs identified through the needs assessment, the total school philosophy, fiscal requirements, and the feasibility of time requirements.

Summative evaluation follows as the program becomes functional:

A fully functioning program should be expected to have:

1. Certified teachers who speak, read, and write the native language.

2. Appropriate curriculum design with materials supporting the community and school educational goals.

3. Ongoing staff development with community training activities.

4. Continuing needs assessment.

5. Teacher commitment.

6. Community understanding and commitment to the program.

In order for data to be meaningful, some criterion or standard must be set up against which it can be measured. Whether the individual project decides to use national norms, a control group, or a percentage of goal attainment set up by the project, it is essential to specifically define some criteria to assess the effectiveness of the program and maintain consistency over a period of time. The summative aspect of evaluation requires analyses of the data collected to determine project impact and the need for modification of program objectives.

Measuring Project Impact

The Federal Rules and Regulations governing the Bilingual Education Act as found in the Federal Register Volume 41, No. 114, Friday, June 11, 1976, (page 23865) specify that "provisions for comparing the performance of participating children on tests of reading skills in English and in the language other than English to be used in the proposed program with an estimate of what such children's performance would have been in the absence of the program" should be made in program evaluations.

There are several methods which may be employed to achieve this comparison. Idealists insist that a control group is essential to measuring effects which may be due to the "treatment" or program. As all educators today know, it is literally

impossible to find control or similar groups to those receiving the benefits of bilingual/bicultural education, because those who need the program are probably already participating in some kind of special program which may include bilingual education. Even where Title VII programs do not exist, there exists Title I programs which provide language development instruction and professionals who speak the second language. Some schools have a form of bilingual instruction where no Title VII funds are being used.

As with the problem of comparing the students' performance with those of a control group, the comparison of students' scores on nationally normed tests with those who comprised the norming sample is a concern. One hears the following comments regarding these norms quite often:

- "Why should we compare our kids to national norms?"
- "The kids in the norms aren't like our kids."
- "They didn't use our kids in the norming sample."
- "They use test items which are not appropriate for our kids."
- "Our kids are on the reservation and have never had the kinds of experiences the kids in the cities have."

So the test publishers have obligingly provided local norms. A problem with this is that the students' scores are being compared to their own. Again, the cry would arise, "But how are our kids doing?"

Some school districts have begun developing local tests. Specific performance objectives are developed and the test, most often referred to as a criterion-referenced test, assesses the

achievement of the objectives.

However, the Federal Rules and Regulations, confusingly enough, further specify there must be "Provisions for reporting pre-test and post-test results on reading tests for all participating children (and, where their performance is compared with the performance of non-participating but similar children, for all such non-participating children) using mean scores, standard deviations, and appropriate tests of statistical significance. No application which fails to include the elements of an evaluation design described in this paragraph will be approved for assistance under this subpart;" Part C (p.23865).

What can the program evaluators do? There are several alternatives. Among these are to compare program students' test scores with those of students who have been in the same schools prior to program implementation, or to compare the program students' scores with those of students in the same district who are not participating in the program.

Unfortunately, students in the "same district" who are not participating in the program probably are not participating because the "need" for the program does not exist. Or, as with many of the Native American programs, there <u>are</u> no other schools in the district.

The authors of this paper feel strongly that evaluation of students' achievement on standardized tests using a pre-test (beginning of the year) and post-test (end of the year) is the <u>least</u> desirable method for assessing student gain. We do not believe that the effects of a bilingual program can be demonstrated after only a six-month period of program implementation. Rather, student progress over at least a two or four year period might be

expected to indicate behaviors due to program influence.

An optimal method of evaluation, therefore, is one of "tracking" students who have been in the program since its initiation. Admittedly, with some school districts, students are "coming and going". However, some students do remain with the program. These are the students whose achievement scores should be analyzed for purposes of longitudinal evaluation. If there is one strong recommendation that can be made, it would be to identify students who have been in the program for several years and compare their achievement test data with those of students with similar characteristics--age, sex, home language--who have not been in the program.

Continuous data gathering must be maintained. It may be that the "true" effects of the program implementation will not be observable until the students are in high school. Dropouts should be expected to decrease and standardized test scores should be average or above if a fully functioning instructional program has been operating.

While the authors feel that longitudinal evaluation of individual students provides the most effective means of measuring the influence of bilingual education, for children of limited English-speaking ability, the guidelines of Title VII do indicate that a measurement of impact must be conducted on an annual basis. Pre-and post-test data must be collected, and comparisons must be made.

Models for comparing student gains have been recommended in a publication, "A Practical Guide to Measuring Project Impact on Student Achievement", Number 1 in a Series of Monographs on Evaluation in Education, distributed by HEW, 1975.

This monograph suggests five evaluation models and provides a procedural guide for model selection. Three of the five models identified are most commonly used. They are analysis of covariance, special regression, and norm-referenced.

A major concern in using the analysis of covariance model is in identification of a control group which has characteristics similar to those of the treatment group. These include such things as living on or near the same Indian reservation and being linguistically and socio-economically similar.

Two special regression models are discussed in the monograph. In both models, "all pupils in a group are pre-tested and those who score above or below a particular score are assigned to the treatment group while the remaining pupils serve as a comparison group." (p.59)

The norm-referenced model compares students' scores to those of students who comprised the norming sample during test construction. This model does not require a comparison group and does require the use of standardized tests.

The most practical consideration is that the tests be administered during the publisher's suggested time frame.

Consultation with an external evaluator taking into consideration unique needs of each program and its population characteristics should determine the evaluation model used.

Federal guidelines require pre-and post-testing and that pre-test data be reported in the interim report. Final analyses using the model selected are provided in the annual report.

PROGRAM COMPONENTS

Performance objectives should be developed for each of the components which reflect the program philosophy and specific activities designed to carry out the intent of the project. As indicated in the previous section, performance objectives must be realistic, measurable and relate to the specific activities that are to be accomplished within the grant period, which is usually defined as one year in length.

Native American Bilingual Projects have special needs which are developmental in nature. These needs must be responded to in both time and effort in order to implement an effective bilingual program. The needs might be identified in a general way as:

1. The need for language development which includes an initial orthography and/or development of a simplified orthography; development of a dictionary, a practical grammar, and material to support literacy in that language.

2. Curriculum development might be identified as the need to define the educational approach based on the language situation of the community, and the development of a specific plan for implementation in the classroom.

3. Staff development which frequently includes training the native speaker in both literacy in the native language and in a teaching career.

4. Community support which includes parents, members of the community, tribal council, and the school community serving the total educational program.

These areas of need might be referred to as the ingredients essential for the implementation of an effective program of bilingual education. They take money and talent to become fully developed. These considerations should be reflected in performance objectives developed by the project. Each of the following sections will discuss the components in more detail.

Material Development and/or Acquisition

The languages of Native American communities require a considerable amount of development prior to their use in an educational program. This development includes writing or revising an orthography or writing system utilizing the English alphabet. If an alphabet or writing system has been developed either by a Bible translator or an anthropologist who had a particular interest in that community, methods used for the development of the alphabet relate to linguistic considerations of the translator.

Most frequently the orthography is developed by writing a symbol for every sound. In most instances, this adds far more symbols than are commonly in use in the English alphabet. To provide an educationally sound approach, it is essential to simplify the system for greater ease of use and better transfer value between the native language symbols and those of the English alphabet.

A curriculum guide or a plan is simply a process of developing an educationally sound approach

for implementation based on the model of bilingual education. This might be a simple transitional approach to bilingual education with its varied types, or it might, in some instances, border on a second language approach. There are various stages in between. Each one is unique, based on a particular language situation and dependent on the desires of those working with the bilingual program. In every instance the plan or program or the curriculum should provide greater educational opportunities for the children that it is to serve. It must be well-planned so that those working with the program totally understand the approach and its educational goals.

For the development of literacy in the native language, there are a number of materials that are essential which must frequently be developed at the project site because of the uniqueness of the language. This might include the development of a dictionary to be available to both students and teachers. A practical grammar book should be available, and reading material in the native language should be developed in quantity and quality to support the efforts to develop reading and writing in the native language.

Other areas such as mathematics, science, music, social studies, etc., might be taught through oral language with some supplemental materials developed by the project staff. If an initial social studies program is to be developed, it is essential for the project to give special consideration to developing the content and perhaps even writing that program in the native language. Most Native American bilingual programs utilize published mathematics programs and teach the content orally through the native language.

Word problems and other supplemental materials are frequently developed by the project. A resource book for music is often developed, usually through translating some of the children's songs to the native language or writing and incorporating other songs from the culture for use in programs. A supplemental science book utilizing a process approach for examining various aspects of the environment may be developed in the native language.

Again, most of the supplemental material identified above will be based on project philosophy and approach. It is the evaluator's responsibility to determine that the component contains performance objectives and that these relate to the program philosophy.

Most Native American programs select already existing English As A Second Language programs. The selected program may be used in its entirety or it may be supplemental with other materials developed by the project. In any case, it is essential that some oral language development take place, utilizing special techniques as they apply to a second language learning situation.

Staff Development

Performance objectives should be developed for this part of the program and respond to the unique needs of the staff members and community. Minimally, these should include:

(1) Literacy training in the native language so that teachers working in the program are able to read and write the native language with a degree of proficiency adequate to support the instructional

component of the program.

(2) Methods of teaching in the bilingual program.

(3) An opportunity for all teachers in the school to gain a basic understanding of the structure of the native language.

All staff training should provide opportunities for career advancement according to individual needs.

Perhaps an area which is most important in Native American Bilingual Programs is that of having the teachers fully literate in both languages. Frequently, native speakers working as paraprofessionals have a negative attitude toward their own language. Some have indicated that they feel the language is too difficult to learn. If this is the case, the same message is conveyed to the children. The teacher who is fully literate in a language will create a rich language environment for the children.

The Summer Institute of Linguistics for Native Americans provides a summer session in Santa Fe, New Mexico each year. It usually continues for a period of eight weeks and students receive eight hours of college credit. This program and similar courses are important resources for developing native language proficiency.

Prior to developing the objectives for an in-service, a needs assessment should be conducted with staff and community members in attendance. Staff development might include courses in basic English, cursive writing,

bilingual theory, classroom management, child development, and techniques of teaching reading. The evaluation of the in-service should be conducted in an ongoing manner with constant feedback for the administrative and instructional staff.

A variety of methods and approaches should be utilized in delivering the staff development program.

A pre-service program, one week prior to the program initiation should be conducted. It should include a session on the theory and purpose of bilingual education. It also should include a complete review of the purpose and objectives of the local bilingual program. It might include sessions on language development, techniques of teaching for language development, and literacy for the native language.

A regular program of in-service, at least one day a month, should be provided following the design based on the needs assessment survey. The purpose of the in-service is to provide a practical orientation to the intended purpose of the bilingual program and to provide for the immediate needs of those working in the bilingual program.

A variety of resource persons should be identified. They should be persons with technical expertise in the areas of linguistic education or other program needs. Local persons should be involved to assist with culture and language development.

A college or university, willing to support the intent of the project, can be extremely valuable in supporting the developmental needs of the project. This might include production of one or more classes on-site which relate to the

immediate needs of the project and provide an opportunity for career advancement. A problem that should be avoided with on-site classes is of overloading the staff with courses that demand work outside the period of time during which the instructor meets with the students. It is important to realize that career advancement is essential, particularly for the paraprofessional.

Staff development should involve not only teachers, administrators, and other members of the school staff, but also members of the community. It is essential to realize that administrators and non-speakers of the native language should be provided an equal opportunity to receive career advancement. In this way the entire team may be able to more effectively integrate the entire program into the school's curriculum.

Each bilingual teacher and paraprofessional should have adequate counseling about his/her educational goals and be provided with a specific plan by which to attain these goals. It is essential that paraprofessionals receive information about accumulated credits and provided with opportunities for summer training. These considerations should be worked out at the local level as they relate to the bilingual program being implemented.

Community Involvement

The development of the Native American Bilingual Program involving a language indigenous to a particular community makes it imperative that there be support from the community at large and the parents of the children involved in

the program. The community is represented by a tribal council which has an extremely important role in the preservation of the language and culture. Thus, it is essential to consult the council about the development and operation of the bilingual program.

A Parent Advisory Committee (PAC) should be formed according to the requirements of Title VII. This committee should meet on a regular basis, usually a minimum of four times a year, and should have an active involvement in planning, implementing, and evaluating the program. The PAC evaluation function might be as simple as coming to school either on a volunteer or assigned basis. The parent may be asked to fill out a questionnaire addressing such questions as, "What did you see? What did you like about the program? What suggestions do you have?"

It is very important in preparing a proposal for bilingual education that the PAC and the public be involved. Title VII requires that there be a public meeting. During the hearing, it is advisable to have a preliminary proposal developed so that parts of the program may be discussed with some degree of understanding.

It is suggested also, that a plan be developed for use by resource personnel. This might include a list of names for different activities to be completed. These persons may be assigned particular days during which they come to the school and perform specified tasks.

A variety of community involvement activities have been identified by other Title VII programs and may be consulted regarding their experiences. Some activities may be effective in one community and not in others.

Some Native American communities have incorporated an advisory committee on culture and the use of language. This committee might represent the council and other members of the community in developing a new orthography. Again, this depends upon the particular community.

Instructional Component

There are a variety of instructional approaches to bilingual education based upon assessment of the language dominance of the children in the community being served. Native American bilingual programs are presently designed as transitional with a variety of types moving to some second language programs. All are designed to facilitate conceptual development through the native language and to provide an opportunity for children to progress more effectively through the educational system leading to a greater proficiency in speaking, reading, and writing the English language. As in each of the other components, it is essential that the performance objectives are designed to meet the needs identified by the needs assessment. These objectives must be specific and measurable.

Before developing a series of objectives, a philosophy of bilingual education must be determined outlining the grade levels which are to participate and the amount of time to be assigned to each language.

In reading the Title VII guidelines, some redundancy seems to occur throughout the section titled "Approach." In spite of this redundancy there seems to be continued emphasis on the fact that there should be a comprehensive approach to the study of the English language and the

dominant language of the child. Thus, it is essential that objectives be established in English and the dominant language, relating to listening, speaking, reading, and writing skills. These objectives should be based on an assessment and diagnosis of the dominant language of the child. If the dominant language is other than English, and the child has a nearly monolingual use of a language other than English, the emphasis should be on the development of an early education program in which most of the content areas are provided through oral instruction in the native language. English should then be addressed as a second language and special English As a Second Language techniques should be incorporated. Where there is dominance in a language other than English and there are trained teachers and materials to support the language, reading should be introduced in the native language. It is essential that there be teachers capable of teaching and materials available to support a reading program in the language other than English, before any attempt is made to introduce such a program.

If teachers and materials are not available, as is often the case in Native American programs, instruction should be initiated orally through the native language as the child enters school, with emphasis on conceptual development, listening and speaking skills, and with some introduction to reading in the native language.

Instruction in the culture and history of the United States is to be emphasized as referred to in the Guidelines of Title VII. This instruction can occur through either language, and may be presented orally. Instruction in the history and culture of the geographic areas serving the children also is to be incorporated as part of the total program, and it too, may be taught either in

oral or written form depending on the age level of the children and in either language.

It is quite obvious that a program of bilingual education for the primary grades which is transitional in nature, with an emphasis on conceptual development in the native language, is quite different from the bilingual education that might be introduced at the intermediate and secondary levels, where the native language is extended and maintained.

Performance objectives should be established for each portion of the program. These objectives should relate back to the needs assessment and they should be specific and measurable to form an important part of the total evaluation plan.

During the site visits provided by the external evaluator, each of the performance objectives should be reviewed and some feedback provided on both the process of collecting data relative to the attainment of the objective and whether or not the objective is consistent with the goals of the program. As the project continues to mature and curriculum guides are developed, there becomes less need for performance objectives within the instructional component, since they are usually constant in the curriculum.

Some projects have made the decision to establish a series of performance objectives for each of the skill areas in both the native and English language. The project then develops a series of tests or checklists to assess the effectiveness of attaining these objectives. These are referred to as criterion-referenced tests.

Management Plan

It is advisable that a management plan be developed which identifies specific project activities, a timeline, and persons responsible. This management chart may be most effectively developed with the assistance of the external evaluator who acts as a group leader with the project staff to define the plan. The plan then may serve as a blueprint for project operation and internal evaluation.

Fiscal control may or may not be the direct responsibility of the Project Director. In either case, it should be clearly understood and defined for effective program management.

Communication is perhaps one of the most important aspects of project management. It may be defined as internal communication which relates to school staff and project staff and external communication which relates to community, parents, school administration, other projects, and the funding agency. School meetings should be held with Title VII staff members to review progress, identify responsibility, and to provide information on school and program development. It is desirable that all staff members have the opportunity to provide feedback on all areas of concern, and that information is provided to the total school staff on request or during staff meetings called by the school principal.

Communication with parents and community members is most effectively accomplished by providing progress reports at tribal council, school board, and other meetings called for the purposes of discussing community interests.

A newsletter might be published monthly which includes pictures of children in the bilingual

program, an adult language lesson in the native language and other areas of interest related to the development of the bilingual program.

District administrators, state departments of education, staff elected officials, and the funding agencies might be included on the mailing list to receive the newsletter and other project information.

Dissemination of information might also include news releases to local newspapers and radio and television stations. The Project Director and staff may assist the community in understanding and accepting the program by making presentations and demonstrations on the use of materials and methods in the bilingual program.

It may seem that this paper dwells principally on program development and implementation, but, in fact, it is impossible to divorce developmental needs from the evaluation. A good evaluation plan, identification of performance objectives and criteria, methodology for data collection, and persons responsible, are necessary to provide regular feedback to staff members, and contribute to measurement of overall program effectiveness.

A good program implementation plan is crucial. From this, and only this, can evaluation take place. Unless one has a fully-functioning program which incorporates the essential ingredients delineated in this paper, evaluation of an empiricle nature cannot take place. Therefore, it behooves the evaluator to assure that goals and objectives related to local needs are being addressed. Then, and only then, can an appropriate evaluation as outlined in evaluation textbooks take place.

Spolsky, Bernard
"Evaluation of Research on Bilingual Education
for American Indians"
A Position Paper, SWCEL
Albuquerque, New Mexico (April, 1970)
ERIC No. Ed 057 093

Dr. George, George P.
Guidelines for Selecting Tests For Use in
Bilingual-Bicultural Education Programs
ERIC Clearinghouse on Language and Linguistics
Washington, D.C. 1975 (ED 108 529)

Pfeiffer, Anita Bradley
Designing a Bilingual Curriculum
ERIC Clearinghouse on Language and Linguistics
1974 (ED 103 181)

Carrillo, Federico Martinez
The Development of a Rationale and Model Program
To Prepare Teachers For
The Bilingual-Bicultural Secondary School Programs
(R & E Research Assoc., 4843 Mission Street
San Francisco, California 94112) 1977.

Scriven, Michael
The Methodology of Evaluation. Rand McNally: in
Stake, Robert E. (ed) AERA
Monograph Series on Evaluation, No. 1
Chicago, 1967

¡ POR LOS NINOS !

Dr. Michael C. Massarotti

Executive Director,
Program Development and Research
4476 West 68th Avenue
Westminister, Colorado 80030

Evaluation is perceived in different ways by different people. Evaluation in this paper will refer, at least, to a systematic procedure for collecting information based on pre-established objectives and subsequently analyzing, reporting and using that information for the purpose of making decisions concerning program development.

The meaningful development of a program rests in part on the ability to answer the following five questions:

1. What do pupils need?
2. Based on that need, where do we want to take them?
3. How do we get them there?
4. Did we get them where we intended?
5. If not, why not?

Inherent in the five questions are the concepts of needs assessment, program objectives, methodology and resources, process impact, system analysis and modification and, finally, recycling.

Many evaluation plans focus only on pupil performance. Basically, a program establishes certain learning expectations as a result of some perceived needs. An instructional strategy is developed which is expected to produce those learning expectations. After the instructional plan is implemented, information is collected on actual pupil performance with respect to what ends were intended. Evaluation in this framework bases decisions primarily on the congruence or discrepancy between expected outcomes and actual performance.

It is important to remember that pupil performance is not independent to some instructional strategy. The understandings and skills of the teacher to deliver the instructional content along with materials utilized are all major parts of the process plan. Meaningful program development cannot occur without systematic consideration of these components and their relationship to pupil performance.

CONSIDERING AN EVALUATION PLAN

Frequently in life the most obvious things are taken for granted, overlooked or are lacking in clarity. Many evaluation plans suffer from this malady. Discussions with various program administrators reveal a lack of understanding on "what" is to be evaluated. An encounter of this kind brings to mind a passage from Alice In Wonderland. You will recall that Alice is lost and in her wandering comes to a fork in the road. In her confusion about which road to take, she spies the Cheshire cat "perched" in a tree. You may recall the dialogue:

> Alice: "Cheshire Puss, would you tell me please, which way I ought to walk from here?"

Cat: "That depends a good deal on where you want to get to."

Alice: "I don't much care where."

Cat: "Then it doesn't matter which way you walk."

A lesson relating to evaluation exists in this passage. It really doesn't matter which evaluation approach is used if "what" is to be evaluated is unclear. At least two steps are crucial as an aid in designating the "what." A needs assessment and clearly defined objectives must precede other steps in program development if we are to meaningfully decide which way we want to "walk."[1]

NEEDS ASSESSMENT

Different types of needs assessment exist in evaluation. Stufflebeam offers four definitions of needs assessment:

1. Discrepancy View - A need is a discrepancy between desired performance and observed or predicted performance.

2. Democratic View - A need is a change desired by a majority of some reference group.

3. Diagnostic View - A need is something whose absence or deficiency proves harmful.

[1]Carroll, Lewis, Alice In Wonderland, New York: Grosset and Dunlap, 1963, pp. 62, 63.

4. Analytic View - A need is the direction in which improvement can be predicted to occur given information about current status.[2]

The discrepancy view of needs assessment appears to be the most popular approach in bilingual assessments.

OBJECTIVES

The importance of writing objectives clearly cannot be over-emphasized. While different kinds of objectives exist, instructional objectives should be stated in such a way that they describe how the learner behaves or will behave. Mager states, "An objective is an intent communicated by a statement describing a proposed change in a learner - a statement of what the learner is to be like when he has successfully completed a learning experience."[3] The lack of clearly defined objectives restricts the efficient evaluation of a program. Remember Alice In Wonderland.

DESIGN

The focus of evaluation can become distorted if the design is not prepared early in the program or established at least before the program

[2]Stufflebeam, Daniel L. Needs Assessment in Evaluation, Paper presented at AERA Evaluation Conference, San Francisco, California, September 22, 1977. p. 2.

[3]Mager, Robert F. Preparing Instructional Objectives, Palo Alto, California: Fearon Publishers, 1962, p. 3.

activities are implemented. The nature of Bilingual Education tends to preclude the experimental manipulation of variables necessary to ascertain the effect of bilingual experiences on pupils. It is common knowledge that cause and effect statements are permissible only from the findings of an experimental study.

Unfortunately, there are instances in which statements of causality are attempted on the basis of data derived from non-experimental studies. It is crucial that decision-makers avoid making this error. The usefulness of the information derived from the systematic evaluation of the programs is restricted typically to development and refinement of processes to produce certain expected learner outcomes. It does not provide a substantive basis for establishing causal relationships.

Some districts indeed are interested in determining the effect of bilingual education on pupil achievement. In this instance, an experiment would be desirable in order to resolve this issue.

Certain characteristics define an experiment. At least an independent variable is manipulated and the pupils are randomly assigned to different levels of that variable. Extraneous variation is controlled to provide assurance that any statistically significant differences may be attributable to the independent variable.

In order to experimentally study the effect of bilingual education on achievement we would, in a classical manner, identify all linguistically different pupils at a particular grade level. Next, we would randomly assign

half to the group (comparison) whose experiences would be identical to the treatment group with the exception of the bilingual component. Care would be taken to insure that the only possible source of variation between the two groups would be the bilingual experience. All other variables such as teacher, classroom, time of day, etc. would be controlled.

The experiment as described in the preceding example is infeasible in a natural classroom setting. Nevertheless, alternatives exist for the district which wishes to make an honest estimate of how well the pupils would have performed without the treatment. Stanley and Campbell[4] offer various quasi-experimental approaches.

In any event it is important to be mindful that regardless of the design, a researcher is obligated and necessarily restricted to draw only those conclusions that are warranted and justified by the procedures and results of our study. In order to yield the type of information an individual is seeking, he must determine in advance the kinds of inferences, if any, which can be derived by his procedures.[5]

Design implications. The design in educational research and evaluation can be an opportunity for truly creative expression. The nature of unique circumstances discovered in every bilingual program will challenge even the most experienced and skilled evaluator. Nevertheless, various common alternative designs for evaluating program efforts exist. Some of these designs

[4]Campbell, Donald T. and Julian Stanley. Experimental and Quasi - Experimental Designs for Research, Chicago: Rand McNally and Co., 1963.

[5]Baker, Robert L. and Richard E. Schuts, Instructional Product Research, New York: American Book Company, 1972, p. 19.

allow greater or lesser credibility regarding possible conclusions that can be drawn or decisions made concerning the relative effectiveness of the bilingual program installed.

The procedural design determines the pattern of information collection. It specifies time arrangements for observations of expected behaviors such as the pre and post-testing of academic abilities, the daily summary of attendance records, the rating of consultants following each workshop, etc. These patterns must correspond to conditions established in the program objectives to be evaluated.

Many projects have established learner objectives that contain the language "improved performance." The manner in which the objective is stated dictates that a pre-treatment reference point should be established. In other words, the concept of improvement suggests that information be gathered on student performance before and after the bilingual experience for that year (pre-post-test observations).

The difference between pre and post-test results can be interpreted as the amount of growth attributable to the educational experience provided for the student. However, the growth cannot be interpreted as the exclusive effect of the bilingual treatment because this program is only a part of the total educational activity for tne students. It becomes necessary then to make some estimate of how much achievement growth would have occurred without the influence or the bilingual activity.

Alternatives. The following is a review of various designs that could be used in determining the relative efficacy of the bilingual component:

1) The optimum design to achieve the stipulated results is a PRE-TEST AND POST-TEST WITH A COMPARISON GROUP.

This model can be most favorable depending on the similarity or comparability between the treatment and comparison groups. Obviously the credibility of the approach is diminished when this condition is not met.

The nature of many bilingual programs precludes the random assignment of students to treatment and comparison groups. If the treatment and comparison groups are composed of intact classrooms, the selection of a comparison class must take into consideration the similarity with the treatment group on selected variables such as proportion of males and females, socio-economic status as measured by the father's educational level, ethnicity and intelligence, to name the most basic.

Because the intact classroom is chosen as a comparison on the basis of selected similarities with the treatment or bilingual group, the approach does not produce identical groups. While this approach minimizes differences, differences between the groups still exist. The following procedure could be utilized to install the design discussed:

1) Identify expertise for evaluation and statistics in district or contract with an external consultant;

2) Establish the evaluation design, including analysis plan;

3) Select tests;

4) Develop a profile of the bilingual class or treatment group on at least:

   a) Average class I.Q.;

   b) Distribution or socio-economic characteristics;

   c) The proportion of males and females;

   d) Ethnic distribution of the students;

   e) Type of classroom organization (self-contained, open, etc.)

5) Identify a school in the district that resembles the school in terms of socio-economic status and ethnic distribution;

6) Select a classroom for the control that is most similar to the bilingual class using the profile established;

7) Pre-test both the treatment and comparison groups at the same time before the onset of the bilingual experiences for the year;

8) Establish a monitoring system during the year to provide assurance that the only basic difference between treatment and comparison groups is the bilingual activity;

9) Post-test both the treatment and comparison group at the same time, usually during May, or as close to the completion of the yearly bilingual activity as possible;

10) Analyze data according to the plan established by the evaluation plan.

2) An alternative design that could be considered is a POST-TEST ONLY WITH A COMPARISON GROUP.

The information provided in the discussion on Design 1 is also appropriate for this design. The basic difference is that no pre-test measurement is administered. Assuming that intact classes were used for treatment and comparison groups, this design provides no information concerning any of the differences that may exist between the groups before the treatment. No statistical adjustment of post-test scores is possible based on pre-treatment differences. Therefore, any post-test differences between the groups may not be the result of the effectiveness of the bilingual activity, but may reflect only those initial differences that already existed before the bilingual treatment. The steps for installing this design would be the same Design 1, with the exclusion of the pre-test step.

3) Another alternative that could be considered would be a PRE-TEST POST-TEST NO COMPARISON GROUP.

While this design is popular in some quarters, it is very restrictive in the types of conclusions that could be drawn from the results.

Since no empirical reference group is included, interpretations concerning the amount of the difference between the pre-test and post-test results attributable

specifically to the treatment lacks scientific precision. Nevertheless, it is possible to estimate the amount of growth expected from no-treatment performance. This estimate could be derived from estimates on typical performance expectations relative to national or local norm group performance. A crude estimate of the treatment effect could be derived from a norm group from the mean difference score of the treatment group. This procedure uses difference scores which pose some specialized problems. The use of difference scores is not recommended or if they are employed, care must be taken to interpret the results within the limitations of this type of score.

The stepwise procedure for using this design would be the same as one excluding any step that involves comparison group activity.

4) A fourth alternative design is the post-test only designed with no comparison group. In certain school districts this may be the best alternative because local conditions and parameters preclude the effective use of one of the first 3 designs. While this design is not as scientific as the alternatives, it is a valid approach that may be used effectively in certain circumstances.

INSTRUMENTS

The selection or development of instruments to collect the information refers to the specific techniques to be utilized in the collection of specific information. Alternative devices include standardized tests,criterion referenced tests, checklists,questionnaires, interview, rating

scales and anecdotal records to name a few. It is imperative that the concepts of validity, reliability and useability be considered for every measure.[6]

Alternative standardized tests. Numerous standardized tests for measuring the performance in reading and comprehension of English exist. Selection of a test must include the consideration of the validity and reliability of the instrument in addition to the cultural relevance of the measure.

The development of a standardized test includes validation information for the prospective user. This information can be located in a technical manual or test manual available from the company producing the test. Further information may be acquired in the Mental Measurements Yearbook by Buros.

While no test reviewed has perfect validity, reliability or cultural balance, the following are selected standardized achievement tests to be considered:

1) The Comprehensive Test of Basic Skills

2) The Metropolitan Achievement Test

3) The Boehm Test of Basic Skills

4) The California Achievement Test

[6]Borg, Walter R. and Meredith D. Gall, Educational Research An Introduction, New York: David McKay Company, Inc., 1963.

Various tests of language dominance are available. For the most part these measures have not been subjected to rigorous assessment of validity and reliability. However, the Bilingual Syntax Measure (Psychological Corporation) and the Test of Language Dominance Assessment (Southwest Research Associates) warrant further investigation.

Extensive annotated lists of various standardized measures may be located in Oakland, Thomas. Psychological and Educational Assessment of Minority Children, New York: Brunner/Mazel Inc., 1977 and Samulda, Ronald. Psychological Testing of American Minorities, New York: Dodd, Mead and Co., 1975.

SELF-CONCEPT

Self-concept, as well as achievement performance are dependent variables of many bilingual programs. A dependent or criterion variable refers to a variable that is measured to determine the effects of a treatment on a population. In both of these instances the target population is students. The evaluation design selected would apply to the collection of information on these student variables.

The nature of self-concept makes it extremely difficult to measure. While it is difficult to identify standardized tests that are valid and reliable, appropriate commercially published tests appear to have more validity and credibility than district developed measures. In the event that instruments are developed locally, professional responsibility must be assumed in making an honest effort to establish the validity and reliability of the measures.

The following are selected commercially produced instruments in suggested order that appear to be the most useful in bilingual programs:

1) Self Observation Scale (S.O.S.)

2) Coopersmith

Recent research in the area of self-concept indicates no agreement on the best measure. With the lack of empirical evidence on the validity of most effective measures, specifically for linguistically different children, the area of measuring self-concept for these students remains unclear. In light of the lack of credulous measures, consideration should be given to the review of the Gordon How I See Myself Scale, Piers-Harris Children's Self-Concept Scale, and Sears Self-Concept Inventory.

Regardless of the instrument selected, it is strongly recommended that continuing research and review of the results produced by the test be conducted at the local level.

PARENTAL INVOLVEMENT

The involvement of parents and community is assuredly a major strength of many bilingual programs.

The need for systematic documentation of parental involvement cannot be overemphasized.

The following are some techniques and procedures that could be employed to document parental participation:

1) attendance logs with actual signatures at all bilingual activities;

2) attendance logs with actual signatures at all scheduled school activities;

3) visitation log with actual signatures to be maintained in all bilingual classrooms;

4) parental questionnaire regarding their involvement in the program;

5) parental interviews regarding the effect of bilingual education on their participation;

6) parent-student diary of our school involvement in bilingual education related activities.

SCHOOL ATTENDANCE

It is recognized that the effectiveness of a bilingual program would affect the school's ability to hold pupils. Historical attendance and dropout information on the target population should be maintained in order to determine whether improvement in attendance in fact has occurred. Further, longitudinal information, on the same group of students, should be collected in order to insure the credibility of conclusions drawn on both aspects of this objective.

The most efficient method of collecting information is direct observation of attendance and dropout behavior. This data can be summarized from the attendance record of each specific class. The information should be collected weekly and reported in order to avoid a monumental time consuming effort at the end of the year.

Additional information on attendance and drop-out could be acquired in the following ways:

1) A parental rating scale, questionnaire or interview, to assess the effectiveness of bilingual education in the areas of pupil attendance and drop-out;

2) Interviews with students to assess their impressions concerning attendance and drop-out.

DATA ANALYSIS

Data analysis is a crucial step in the evaluation process. This step includes the plan for processing the results of an observation, interview, test, or questionnaire. It is process of data reduction. Depending on the nature of the evaluation question to be answered, or objective to be evaluated, data analysis can range from a simple descriptive statistic such as a relative frequency to a more sophisticated inferential statistic such as an F-ratio.

Statistical analysis is in effect the process of gathering, organizing and interpreting numerical data. Descriptive statistics involves the description of a particular group. Inferential statistics, on the other hand, leads to a judgment about a whole population to which the sample investigated is presumed to be related.[7]

John Best attempted to put the use of statistics in proper perspective. He said, "Statistics is the servant, not the master, of logic; a means, not an end, of research. Unless basic assumptions are valid, unless the right data is care-

[7]Best, John. Research In Education, Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1970 p. 259

fully gathered, recorded, and tabulated, and unless the analysis and interpretation are logical, statistics can make no contribution in the search for truth."[8]

POSTSCRIPT

No two bilingual programs are identical. While similarities are common, great caution must be exercised in "transplanting" someone else's plan. The identification of those variables that make a program unique is essential. With this perspective, the development of an evaluation scheme becomes a creative and challenging act. It would be wise to solicit the assistance of an individual who has demonstrated expertise.

Finally, when all is said and done, regardless of the sophistication of the evaluation plan, if evaluation results do not impact program development at the learner level then the time, effort and money expended has been accomplished for someone or something other than the pupil. The ultimate payoff of a good evaluation effort is improved learning opportunities for kids!

[8]Ibid, p. 259

BIBLIOGRAPHY AND SELECTED REFERENCES

Baker, Robert L. and Richard E. Schutz. Instructional Product Research, New York: American Book Company, 1972.

Best, John. Research In Education, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

Borg, Walter R. and Meredith D. Gall Educational Research An Introduction, New York: David McKay Company, Inc., 1963.

Campbell, Donald T. and Julian Stanley. Experimental and Quasi-Experimental Designs for Research, Chicago: Rand McNally and Co., 1963.

Carroll, Lewis. Alice In Wonderland, New York: Grossett and Dunlap, 1963.

Kerlinger, Fred N. Foundations of Behavioral Research, New York: Holt, Rhinehart and Winston, Inc., 1973.

Mager, Robert F. Preparing Instructional Objectives, Palo Alto, California: Fearon Publishers, 1962.

Nunnally, Jim C. Psychometric Theory, New York: McGraw-Hill Book Co., 1967.

Oakland, Thomas, et al. Psychological and Educational Assessment of Minority Children, New York: Bruner/Mazel Publisher, 1977.

Samuda, Ronald J. Psychological Testing of American Minorities, New York: Dodd, Mead and Co., 1975.

Sanders, James R. and Dean H. Nafziger. A Basis for Determining the Adequacy of Evaluation Designs, Kalamazoo, Michigan: Evaluation Center, College of Education, Western Michigan University, 1976.

Stanley, Julian C. and Kenneth D. Hopkins. Educational and Psychological Measurement and Evaluation, New Jersey: Prentice-Hall, Inc., 1972.

Stufflebeam, Daniel L., et al. Educational Evaluation and Decision Making, Itasca, Illinois: F. E. Peacock Publishers, Inc., 1971.

Stufflebeam, Daniel L. Needs Assessment In Evaluation, Paper presented at AERA Conference, San Francisco, California, September 22, 1971, p. 2.

Worthen, Blaine R. and James R. Sanders. Educational Evaluation Theory and Practice, Worthington, Ohio: Charles A. Jones Publishing Co., 1973.

EVALUATION OF TITLE VII PROGRAMS OVERVIEW AND DESIGN

Dr. Charles F. Leyba

Director, Dissemination and Assessment
Center for Bilingual Education
California State University
5151 State University Drive
Los Angeles, California 90032

Overview

Because the Title VII bilingual program has recently come under scrutiny and criticism basically for projects' lack of evidence, statistically analyzed, as to their effectiveness in teaching children basic skills in English[1], it is important to attend to this in the light of some historical realities.

At the time of the first funding of bilingual projects, the then Associate Commissioner of the Bureau of Elementary and Secondary Education, Dr. Leon Lessinger[2] attempted to put the bilingual program on an accountability basis. In essence this meant that strictly formulated performance objectives were to be designed for each project. Project director management,

[1]Cf. Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program, Vol. III, Year Two Impact Data, Educational Process, and In-Depth Analyses, American Institutes for Research, Jan. 1978.

[2]Dr. Lessinger authored a seminal text explaining the accountability strategy entitled, "Every Kid a Winner," Simon and Schuster, 1970.

together with classroom teaching strategies were to be focused as concentratedly as possible on the achievement of the objectives.

To aid in achieving accountability in management, conferences were held at which experts on this topic provided training for the first bilingual project directors and Federal Program Officers. This occurred in the months preceding the 1969 academic year. The majority of the projects then, as now, served the largest non-English speaking minority in the United States, the Spanish speakers, particularly those in the Southwest.

At the same time, the revolt of the black minority in the inner cities of large metropolitan areas had spread to other minorities, especially the Mexican Americans. This rising tide of resentment at exclusion from participation in mainstream American life centered largely on the schools and their academic programs.

In effect, the schools were criticized under the aegis of the now trite, but still meaningful word "irrelevant". The academic content did not reflect the culture and language of the minorities and the school staff, at every level, contained few persons representing the minorities.

The major stress of the student movement was to reform the curriculum and admissions procedures and secure (where applicable) student financial support. At the elementary school level--this is where the Title VII efforts were largely concentrated then and now--it meant bringing the community into the classroom. The community was largely represented by the Chicano movement, as far as Mexican Americans were concerned, and we shall concentrate on this minority from here on in the paper. As far as education is concerned, the Chicano movement represented a frank, unabashed

irredentism, a demand for a return to the origins basically of a disenfranchised (here and in Mexico) group marked by poverty and under-education.

As a concomitant, the movement rejected, as far as possible, all middle class (gabacho) styles of life. To the extent that the schools were concerned, this meant among many other things a rejection of any attempt to measure performance of children in the classrooms by instruments that were generated by and for that life style. There were, of course, no objective, standardized testing measures to determine cognitive growth in children using their home language.

Because of this lack and, in many respects, target constituency pressure, the projects became a property of the community so that evaluation of early bilingual projects was not of the hard, objective sort.[3]

The second historical fact was that, even if measures of bilingual student achievement had existed, trained personnel to staff the project were not available. As is noted by Dr. Jose Cardenas:

> In general, bilingual education programs have failed to identify, recruit, employ or train the staff necessary to implement a bilingual education program. In one major city with an extensive enrollment in bilingual education programs, 92 percent of the staff in the programs were monolingual English speakers.[4]

[3]In 1975, AIR published The Identification and Description of Exemplary Bilingual Education Programs in which out of 175 bilingual projects investigated, seven were found to have acceptable objective measures of outcomes.

[4]Language, Ethnicity, and the Schools, Epstein, N., Institute for Educational Leadership, George Washington University, Response I, p. 73, Aug. 1977.

In effect, then, Spanish bilingual education projects were staffed in large part by monolingual English speaking teachers and a Spanish speaking aide. These latter were frequently students in the lower division undergraduate college with no training in basic pedagogical methods. Moreover, they were often restricted in the number of paid hours they would serve in the classroom to avoid being afforded fringe benefits such as medical insurance. It was not until the 1974 Bilingual Education Act that formal arrangements were made for teacher training.

Clearly, these and other realities had profound effects on the output and evaluation of bilingual projects.

Evaluation Design

Evaluation of a bilingual project can proceed along two paths: criterion-referenced testing (CRT) and standardized testing.

Criterion-Referenced Tests

Generally speaking, CRTs are locally developed tests[5] and optimally should be administered at the beginning of the school year and at its end, though they may and should also be used over shorter time spans as pre- and post-test measures for teaching units during the academic year.

[5]The Spanish Curricula Development Center has produced core materials together with widely field-tested CRTs for each of the major Hispanic groups: Puerto Rican, Cuban, and Mexican-American students.

Usually CRTs are most suitable for measuring student outcomes relative to subject matter of such a local or regional variety as to have little counterpart in standardized measures. Also, they can be developed to determine pupil growth in the home language particularly in the early stages of a project when the language (Spanish) reflects a quantity of terms known to be regional in nature. Certainly, one distinct value they have, in addition to those mentioned, is that they can be designed to contain a larger number of items to more accurately indicate pupil understanding of a concept or operation. For example, a standardized test may include only two or three items to measure student understanding of a single operation, e.g., the addition of one type of fractions. A criterion-referenced test, as teacher made tests also do, can contain multiple items for determining the dimensions of student knowledge. As a result, they are also excellent diagnostic/prescriptive measures. Perhaps the ideal use of CRTs is in both summative and formative evaluation, with briefer CRTs used as distinct subject matter divisions during the school year (formative) and as pre- and post-tests at the beginning and end of the school year (formative and summative).

The development of CRTs presents usual but not prohibitive problems for the bilingual director and his staff. Some of these are listed below:

1) The curriculum. It is not an exaggeration to say that the course of study in a given content area is usually laid out for the teacher by the text that is being used. The chapter headings with their own subdivisions can easily be divided to present a semester or year's instruction. For the more veteran teacher, the course of study is a set of ideas in the teacher's head to be covered over a certain period of time.

The net result is that, the teacher is forced to sit down and plan out what basic facts, concepts and operations are to become the responsibility of the instructional program and to place them in sequence according to the teacher's or district's philosophy.

In order to develop CRTs, the teacher must be able to analyze subject matter into the above components and answer the question, "What am I making myself responsible for as far as what my students will know by year's end?" A step requiring more sophistication will require determining what order or sequence will be given to the facts, concepts, and operations. At this point the teaching philosophy comes into play.

2) The next step involves writing the facts, concepts, and operations into performance objectives. There are many fine authors that can be referred to in this respect. Popham and Mager are among the leading.[6] Because they and others have developed in detailed fashion the characteristics of genuine performance objectives and the steps for developing them, attention here will be given to other factors.

Besides the basic necessity of writing an objective that can be measured in observable terms[7], it is necessary to decide if what you think your objective is, is in fact what it really is. One bilingual project, for example, administered a physical education test in Spanish. The objective required that the students perform certain acts showing some agility. How does a student

[6]Criterion-Referenced Measurement, an Introduction, Popham, W. James, Educational Technology Publications, 1971.

[7]Every author of texts regarding performance objective treats this topic at length.

pass this test? By being agile or by understanding commands in Spanish? What is the objective? Besides being confusing, this test has to be administered individually - a long, debilitating process for teachers and students.

Another example has to do with following commands, taken again from a Title VII evaluation. The tester asks the student to count objects on a table. The child follows the direction but cannot count in proper sequence. What is the objective? That the child understands the command or that the child can count in sequence?

Obviously then, the development of performance objectives requires clarity and above all patience in going through all the revisions that almost inevitably are required.

3) Another characteristic or need is that the objective be one for which the teacher is genuinely responsible. That is, the objective must be one which the instructional program guarantees the student will come to know over a period of time. Put in another fashion, teachers frequently don't know the extent of the information the student brings to the classroom. In this case then, teachers may begin the instructional process by teaching what the students already know. This in itself is not entirely undesirable since one of the fundamental goals of any educational process is building confidence in the learner. This should involve the opportunity to respond easily and correctly to questions on the basis of previously acquired knowledge.

However, extensive experience in constructing criterion-referenced tests indicates that teachers tend to peg objectives at a level that allows for little growth on the part of the learner and, consequently, little teaching effort.

Again, an example will best illustrate the point. A criterion-referenced sub-test was developed, different from the one referred to above but dealing also with commands given in Spanish. The sub-test contains seven items and was administered at the beginning, middle and end of the year. The first two administrations were for formative purposes. The results are listed in the following table:

Table 1

Percentage of Children Responding Correctly

(N=27)

Pre-test	85%
Mid year	93%
Post-test	100%

Clearly, the teaching objective which this test was designed to measure was intended to elicit or teach behavior already known by the children and for which the school can claim no success or responsibility. Moreover, since each child represents roughly four percentage points, it is clear that the teacher had only to instruct four students over nine months how to respond correctly. This is hardly an objective requiring or deserving an entire academic year's effort.

One discouraging result of developing objectives that assess knowledge attained elsewhere is that little progress can be recorded at post-test because of the very high pre-test scores. Teachers in such a situation once remarked, "We teach alot more in the bilingual program than this subtest measures." That being the case, it is necessary to ask them to specify precisely what this

"alot more" is that they reach. In other words, one must return to the task of rewriting objectives. At this point, unless extraordinary good will and high morale are present, the breakdown of staff cohesion and teacher sense of effectiveness begins, with all the attendant circumstances of blaming federal funding, criticizing the Director, complaining of having to carry out two instructional programs rather than one, etc.

A more optimistic example is the following, also taken from a Title VII project.[8] The objective here is that the child correctly name selected objects found in the classroom and parts of the body. The number of items in this sub-test, delivered in Spanish, is 21.

This table demonstrates an objective that leaves a 62% stretch of learning to be completed over an academic year's time. It requires long-range planning and effort. It is an objective teachers can be proud of.

Table 2

Percentage of Children Responding Correctly

(N=27)

Pre-test	38%
Mid year	72%
Post-test	94%

Taking the above into consideration, a set of criterion-referenced tests should, perhaps, be developed to measure the attainment of short, intermediate, and long-range objectives. The criterion for determining the differences among these

[8]All the Title VII examples in this chapter are from projects in a large metropolitan area.

objectives is the percentage of completion or percentage of correct responses at pre-test time. For example:

Table 3

Type of Objective	Percentage of Correct Responses (pre-test time)
Short-range	65 - 80%
Intermediate	45 - 64%
Long-range	0 - 44%

Criterion-referenced tests should, as noted, be composed of a combination of the above. A possible combination is the following: 20% short-range, 30% intermediate, and 50% long-range objectives.

Typically, however, at first most sub-tests will show an average correct response of over seventy percent. The project director who proceeds to develop, together with his staff, a set of criterion-referenced objectives and associated test items should commit two years to their development. All testing during that time should be to find out which objectives to retain and which to discard.

Finally, there should be two comparable or matched sets of criterion-referenced tests, one for use at the beginning of the year or period and another at the end of the year or period. Otherwise, children become test wise and can respond accurately to items simply by having learned the desired answers in the previous testing period.

Until such time as the criterion-referenced tests are fully developed, the reporting system need only give percentage of correct responses to the items. Later, more sophisticated analyses can be utilized. At that time also, control groups can begin to be employed. Selection of control groups is described in the following section.

Standardized Testing

(This discussion will assume the use of a standardized test administered in English since there is relatively little experience with, or wide acceptance of such tests administered in another language. Moreover, if current information is correct, forthcoming legislation for Title VII will aggressively require the demonstration of the early acquisition of English language skills by the project participants. Standardized tests administered in English are singularly useful for this purpose despite the fact that they will continue to be biased assessment tools for the partial bilinguals.)

Standardized testing, in fact all testing, has a function analogous to holding up a mirror before a person. A test reflects what is there and what is not there. Standardized testing then, is an issue subjunctival to or dependent on a program. The real and substantive question deals with the extent to which a given program provides for and conveys content to students which is commonly found in the minds of students who were used to establish norms for the examination. Once one is assured of this condition, any one of the commonly used standardized tests can be used to assess student growth.

Of course, another substantive question deals with the language by which that commonly known content will be purveyed to the student. Over time if bilingual programs provide growing capability in English language skills while simultaneously developing content knowledge, the use of standardized tests administered in English should be less an issue.

Current federal legislation requires the following of Title VII projects:

a) Instruction using English as a teaching medium and the study of English itself.

b) Instruction in the child's native language to the extent needed to progress effectively through the educational system.

c) Instruction in the appreciation of a child's cultural heritage to the extent needed to progress effectively through the educational system.

d) Demonstration of effective ways of developing competence in English language skills in children of limited English-speaking ability.

The standardized test administered in English will satisfy requirements a & d above. (Criterion-referenced tests can be used to measure c. For item b there are many tests to measure Spanish language ability such as the Inter-American Series Spanish version.)

Besides the above reasons for standardized testing in English, there is the very real need to prove to many in the target and non-target constituency that students in the Title VII program make comparable progress in what may be called the standard school program usually taught in English.

1) There must be comparison. This can be accomplished in a variety of ways.

a) The most common technique is to establish a control group intended to be compared with the treatment or bilingual group.

b) A comparison can be made between the scores achieved by the bilingual group and the norms provided by the testing instrument manual.

c) Finally, a comparison can be made between the scores attained by the bilingual group and scores achieved by pupils in the standard school program in years prior to the introduction of the bilingual program.

Of the above, we will attend to a & b since c is really a variety of a. Item c above applies to situations where the bilingual group enrolls all the students leaving no pool from which to select a synchronic control group. It usually applies in small school districts and isolated areas.

In the following discussion, a & b will be spoken of together.

Ideally, the participants selected for the bilingual and control groups should be randomly selected. Otherwise the effects or outcomes in scores may be the result of characteristics in the participants rather than the treatment they are given. An example many bilingual programs have experienced is of middle class immigrant students from Mexico outperforming fellow students in the bilingual program for obvious reasons. Federal regulations, of course, require that parental consent be obtained for participation in the project. This has the effect of having students in the program whose parents have an active interest in bilingual education and whose children may be additionally motivated to succeed. Because of

parental consent to enter the project, successful outcomes for bilingual groups measured in English version standardized tests have been attributed by researchers to the fact that the parents enroll their dominant English ability students in the project so that they can learn spanish.

The effect of this required selection also introduces bias in the control group since, if you select the students motivated by their parents for the bilingual group, you have the effect of reducing the number of such students in the control group. Where there are very large numbers to select from, this should have minimal effect.

In selecting students for the control group, a general pool of candidates should be identified who come from families closely matched with families in the bilingual group as to level of income, level of educational attainment and number of siblings. From this pool a random selection of candidates for the control group should be made.

For reasons associated with t tests of significance, the numbers of participants in the comparison groups should be 30 or more for each. Moreover several classes for each group is also desirable since the effects of outstanding teachers can be submerged in the larger means or averages.

Analysis of Scores

The most common style of reporting standardized test scores is to report them as pre-test and post-test means and differences for both the

bilingual and control group as indicated below. The example is for the end-of-year report.

	Bilingual	Control
pre-test	39.4	40.8
post-test	52.2	43.7
difference	12.8	2.9

The pre-test scores are the mean score achieved by the two groups. At the time the pre-test is administered, the names of those students who missed the test should be noted and, should they take the post-test, their scores should be excluded in calculating the means and vice versa. Only the scores of students who took both pre- and post-test should be incorporated in calculating the means.

As noted in the example above, there is very little difference between the means at pre-test time. This is a desirable condition since, on the face of it, the groups are evenly matched at the beginning[9] indicating the possibility that larger differences at post-test favoring the bilingual group are attributable to bilingual instruction.

In the example above, there is a large difference between the means at post-test giving rise to the sense that this difference has some significance. Size itself is only an indication and not an unmistakable one at that. Therefore, a statis-

[9]This is an additional and important use for standardized tests, namely, to determine how closely matched the treatment and control groups are in the beginning.

tical measure of significance must be employed. Such a measure of significance is used to determine whether the size of the difference occurred by chance. (Human life, outside of specially designated areas such as race tracks and casinos, can be interpreted as a systematic attempt to destroy chance.) This is why significance is expressed in numbers such as .05, .01, etc., meaning that there are 5 or 1 chance out of 100 that the large difference occurred accidentally. Differences have no real meaning unless they are statistically treated.

In the example given above, the difference between pre- and post-test is significant at the .05 level favoring the bilingual group. A t test was used to determine the level or degree of significance. The evaluator, director and teachers therefore, can be reasonably sure that in 95 cases out of 100, the large difference (12.8) favoring the bilingual group is due to the program.

Group achievement can also be reported through grade equivalences which are another way of expressing raw scores. (At the end of the chapter, grade equivalences are displayed in graph form. See Figures 1 and 2.)

The foregoing pages have attempted to put forth a straightforward evaluative approach for basic programs in bilingual education. There are other approaches and variations that are used depending on the training of the personnel involved.

In the following paragraphs, the results of such an approach in the evaluation of the Title VII project (grades 1-6) are displayed.[10]

[10]Grade One is not included since the MAT does not give grade equivalence for that grade.

The two graphs show results in Total Reading and Total Math - the only sub-tests found throughout the six years during which the Metropolitan Achievement Test (MAT) was used as the measurement device.

Four groups are reported on:

1) The national norm

2) The longitudinal bilingual group (LBG)

3) The non-longitudinal bilingual group (NLBG)

4) The control group

The longitudinal bilingual group and the non-longitudinal bilingual group numbered about 65 children in 3 grades at 3 schools throughout the 6 years. The difference between the LBG and NLBG is that the former group was composed of children continuously, uninterruptedly enrolled in the program, while the latter was composed of children brought in to fill the positions made vacant through departure. Thus over the 6 years the NLBG had children who were in the bilingual program three or four years but never on a continuous, uninterrupted basis.

The control group was made of 25 children to match the grade level of the program in any given year. Nine were selected from one school, eight from each of the two others. It was assumed that, because these children came from the same attendance area served by the three schools, they were reasonably matched with the bilingual group by SES and family income.

Conclusions

Conclusions drawn from the findings are as follows:

1) Long term (of graphs) continuous bilingual education produces strong positive results in cognition as evidenced by the following:

a) The longitudinal bilingual group surpasses at post-test all local groups (non-longitudinal and control) in nine out of ten comparisons in both Total Reading and Total Math.

b) The longitudinal group alone surpasses or equals national norms.

2) Short term and/or interrupted bilingual education produces positive results in cognition as evidenced by the following:

a) The non-longtudinal bilingual group surpasses at post-test the control group in seven out of ten comparisons in both Total Reading and Total Math.

3) For comparable groups undergoing bilingual treatment continuously, transition to full instruction in English may optimally occur during grade five for Total Reading and during the fourth grade for Total Math.

Post Script

Any basic program in bilingual education is primarily a set of services provided to children and not a strictly scientific experiment. As a result, any evaluation of such a program is going to leave many questions unanswered and raise others. This is as it should be.

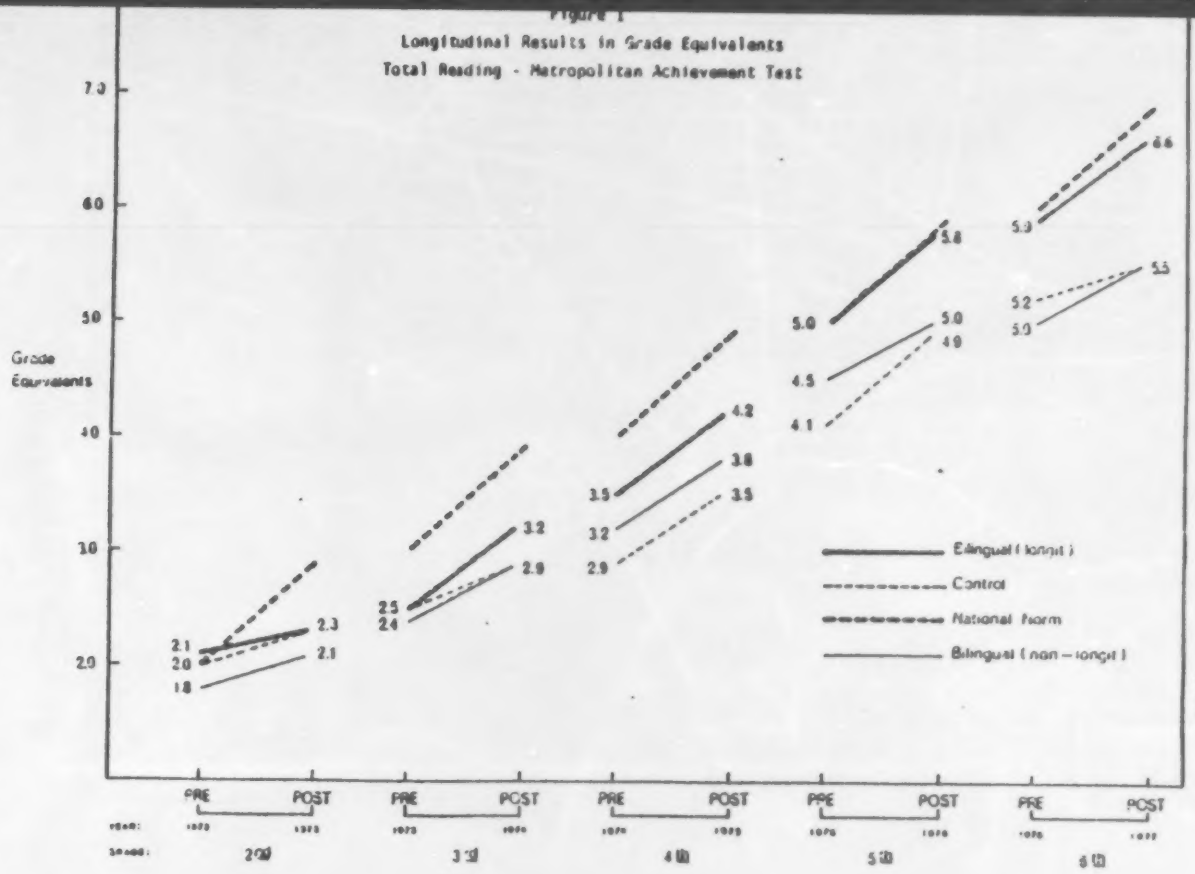
The main point is to take reasonable, straightforward steps to evaluate a program by the most effective means possible. The examples given should only be taken as guides.

Over time, if reasonable steps are taken, evidence will accumulate to demonstrate the desirability of bilingual education.

Figure 1

Longitudinal Results in Grade Equivalents

Total Reading - Metropolitan Achievement Test



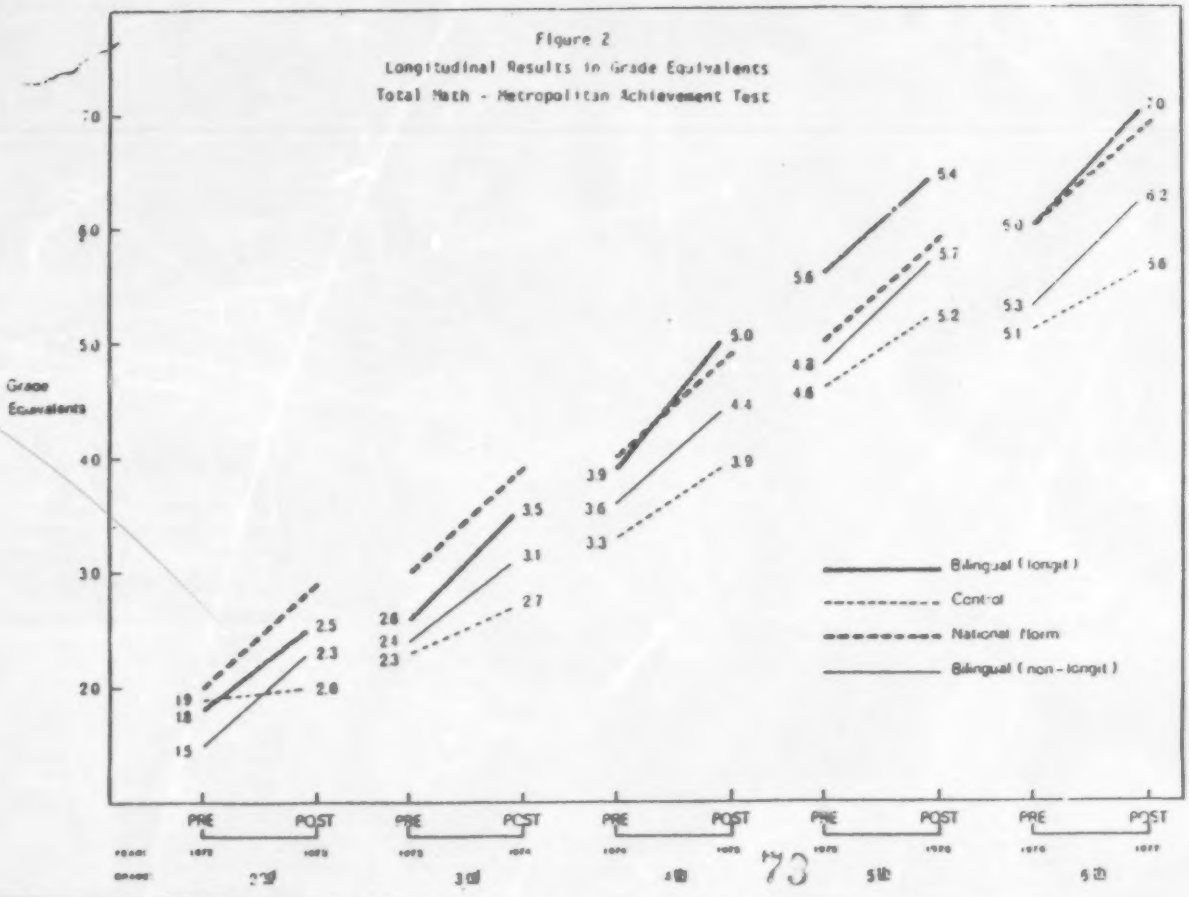


Figure 2
Longitudinal Results in Grade Equivalents
Total Math - Metropolitan Achievement Test
Grade Equivalents
70
60
50
40
30
20
PRE
POST
2nd
3rd
4th
5th
6th
19
18
15
25
23
20
26
24
23
35
31
27
39
36
33
50
44
39
56
48
54
57
52
50
53
51
62
56
Bilingual (longit)
Control
National Norm
Bilingual (non-longit)

EVALUATING BILINGUAL EDUCATION PROGRAMS WHILE REMAINING COHERENT

Dr. Robert T. Reeback
Program Evaluator
3029 Delano Place, N.E.
Albuquerque, New Mexico 87106

ROBERT T. REEBACK is a specialist in language assessment with the Cultural Awareness Center Trilingual Institute (CACTI) in Albuquerque. Dr. Reeback was Associate Director of the Southwestern Cooperative Educational Laboratory, Inc. from 1969 to 1972, and has served as a consultant, program auditor, or evaluator for numerous educational projects in the Southwest. He occasionally teaches linguistics courses at the University of New Mexico. This past year he edited El Noticiero, the newsletter of the New Mexico Association for Bilingual Education.

Introduction

This paper is concerned with living with evaluation requirements. It is not a how-to-evaluate manual, although it offers some advice for bilingual education project directors. My bias is that the most important evaluation activity is seeking information that will help a local project director and staff to conduct an effective project. Some effective projects can yield convincing data. It is distracting to try to evaluate a project and a national program at the same time. Compiling evaluation results for accounting to the government or for appealing to public opinion does not necessarily lead to effective projects.

The main topics in the paper are (1) a critique of Title VII regulations concerning evaluation, (2) some examples of misguided or inadequate evaluation, (3) a review of some persuasive evaluation results with an eye toward what made them convincing, and (4) suggestions for reconstituting evaluation in the context of Title VII and adjusting to it in the meantime.

Local projects and national policy.

After about ten years of bilingual education projects, one finding that emerges is that clear general findings typically do not emerge from local project data. Evaluating any single local project has to be distinguished from more general assessment of the wisdom, acceptability, or efficacy of our national policy concerning bilingual education. Nationally we appear to have accepted a transitional view of using minority languages for instruction, specifically, the "other" language is used as a bridge to improving English. The Title VII ESEA logic is compensatory; the local projects are expected to remove language barriers to academic achievement and permit children whose English is limited to catch up.

Various constituencies offer competing views. In some places the persons who plan, staff, and evaluate the local version of bilingual education may be striving to maintain two languages, and may not accept a transitional, compensatory program as bilingual education.[1] Irrespective of competing views, the Title VII evaluation requirements carry out the transitional intent of the legislation. Projects are funded to enable certain children to advance through school; their evaluation reports must attempt to establish whether or not that advance takes place. Although the logic may be misguided, project directors live with their

evaluation responsibilities. It is understandable that project directors tend to identify the required comparisons of achievement--and especially reporting them to Washington--with all of evaluation.

The local project is the focus of budgeting, staffing, curricular, and other decisions that every project director is aware of. The directors are well placed for evaluating their own project or overseeing its evaluation. Especially because they and their staffs have to make decisions about that project, they are in a position to seek and use information that will help with the decisions. Here is an example: Should we be teaching mathematics first in the other language? One can consult research reports or experts for advice. One can confer with one's own staff. One can examine the project itself for clues. What clues? For instance, what kinds of math achievement do we attain now after two years, after four years? How are we really introducing mathematics right now? Are our teachers qualified for and comfortable with teaching math in the vernacular? Do parents have strong expectations about this? Systematic efforts--not necessarily using paper and pencil--to examine the project in these ways are evaluative, and the project directors can utilize evaluation resources for such efforts too.

The local project directors are not nearly so well poised for abstracting generalizations from numerous projects and types of projects. It is not clear if anyone is well placed for that task. We have had nationally contracted studies of "process" and "impact," and interest in "planned variation."[2] We have research studies

and reviews of many studies, such as Engle's (1975). Some of these are very helpful. I happen to think that assessments of the entire bilingual education effort are extremely difficult to orchestrate nationally, and that one needs a very long perspective in time to arrive at judgments. In any event, it is not the local project directors' job to compile information that will allow decisions about the merits of bilingual education in the abstract, or even of the federal program of efforts involving bilingual education.

Accountability and decision making.

A contract between the demands of accountability and those of decision making underlies this discussion, in addition to the distinction just drawn between local project management and central policy making. The accountability-decision making contrast has often been raised;[3] both purposes of evaluation have to be met at both the local and national levels. Briefly, accountability involves the retrospective defense of actions taken, for example: how the money was spent, how the staff was deployed, how the objectives were set, what objectives were met. Decision-making requires information in advance: what plan is most promising, which objectives should have priority, what kind of staff is needed, how should our approach be modified next year, and so on.

Public opinion.

Besides reports on what one did with the money (accountability) and assessments used to calibrate the project (decision making) there is another function of evaluation labeled here "appealing to public opinion." A northern New Mexico school district, for example, reports to its constituents

and in the newspaper that its long established bilingual education program is a success: Nagging questions have been resolved. Bilingual education does not hamper, it boosts achievement. The report cites the evaluation proper: on criterion-referenced tests of social studies, sciences and other topics, the project students register greater pre/post gains than do the non-project students; on some tests of English the gains are roughly the same. Then we get some amplification: the project has grown five-fold since it began; the number of separate schools has tripled; the size of the staff has quadrupled and will have to increase further. Not only that, but principals who used to be skeptical have had their doubts "quelled." Clearly this is a successful project.

Maybe the project is a success: I believe it is a pretty good project. What I am intrigued by is the function of the evaluation. Roughly, the report is designed to discourage dissent and encourage support in the district staff and community (many of whom are the same people). The evaluation results, proper, are not that convincing, at least not without more background. For example, we don't know how important the objectives on the criterion referenced tests are, or how searching the items are. We don't know how students are placed in the project classes or into "control" classes, or how staff is deployed in and out of the project. We know nothing about variability within groups. For this appeal to public opinion these gaps do not matter in the way that, say, the budget and staffing statistics matter. The appearance of favorable evaluation results is more crucial than the reality.

The federal bilingual education apparatus is a similar position; congressional and other opinion can be appealed to with apparent positive (or negative) results, as we have seen in the case of the AIR national evaluation of Title VII programs. When a federal official calls for "success stories", he wishes to counter "failure" or "waste" stories coming from critics. The emphasis is easily displaced from success to stories. In search of ammunition for persuasion, the Office of Education is tempted to increase the rigor of its requirements pertaining to evaluation, in the name of accountability.

Regulating evaluation.

The Title VII regulations concerning evaluation feature obligatory comparisions. A local project must compare reading achievement in the project with what that achievement might have been without the project. A local project must compare pre- and post-measures of reading achievement. Just to be eligible for funding, a project has to appear as if its students will progress through school at the same rate as other students. Some obligatory comparisions are discussed below. I view these regulations as tending to charge the local project directors with just the task I believe should not be assigned to them, namely, informing national policy decisions. It is as if the Office of Education wished to make local evaluations so uniform that their results could be aggregated in box-score fashion. The AIR report exhibited aspects of this approach. Of course, any results from the projects might influence it more. Nevertheless, strict regulatory controls over evaluation design could tend to decrease the possibility that truly persuasive results will emerge from local projects. The outcome instead could be more short term concentration upon narrow effects, more diversion of

attention from any evaluation that might support local project management, and more stress on the unilluminating horse race aspects of implementing bilingual education.

Restricted range of effects.

The Title VII regulations, for example, require an assessment of how the reading achievement of students in the bilingual education project differs from what could have been expected of them without the project. This is a relatively narrow set of effects upon which to focus evaluation. A wider range of effects, just briefly, might include changes in willingness to use one or both the languages in question, changes in later schooling or career choices, realignments of school and community constituencies, more use of teachers as counselors, changes in disciplinary incidents, more parent interest in literacy, and many others. Unfortunately, requiring Title VII projects to produce the data on individual reading achievement works to preclude investigation of other kinds of effects, because of economics, and because directors and others respond to evaluation regulations as to other regulations--by minimally fulfilling the requirement. While the regulations do not prohibit searching for other effects, it is unrealistic to expect local project staffs to do so.

The examples just given are an elaboration of Mackey's (1977) categories of effects of any bilingual education program, namely (i) effects on the language repertoire of a community, (ii) other effects on the community, and (iii) effects on individuals.[4] The Title VII regulations direct evaluation toward one type of measurement of the last type of effect, (here, achievement scores of individuals in reading). Ironically that is the

one kind of effect that is already overstudied; raising reading scores is a projected outcome of numerous special programs in education. If one must compare bilingual education projects horse-race fashion with whatever else happens to be offered, this choice of effect is convenient. It is roughly like estimating the worthiness of all persons by how much money they make. Macdonald (1974) treated some of the philosophical liabilities of an evaluation that is focused on such a narrow range of effects. His discussion was concerned mainly with the stultifying impact of achievement score increases and similar "preordained" objectives on education; those who evaluate are not open to important growth of individuals, changes in priorities, or opportunities for productive investment of energy.

Without appealing to any esoteric potential effects, there are other kinds of information that might help a director to strengthen a project. For example, reading score gains will not show what students do while they are in the bilingual education project; whether or not they seek to read; whether or not they put effort into writing; whether or not they are more vocal about their demands; pose more or better questions to a wider audience of adults. Getting data about all these things is every teacher's task. Assessing such behavior of bilingual education participants intermittently could help a project director make such adjustments as increasing (or decreasing) emphasis on other language writing, suggesting new types of interest centers. And the effects just listed do not even mention specifically mathematics, science, social studies, music, art and other areas in which bilingual instruction could show effects, nor secondary or tertiary effects, such as changes in the eagerness of teachers to work in the project.

Some other kinds of project evaluation that compete for attention with the obligatory reading comparisons are formative studies of materials that the project staff devised or wants to try out, surveys of which kinds of staff training should get the highest priority or proved most valuable, assessments of staff attitudes about such issues as the urgency of meeting requirements for bilingual endorsement, or the anticipated impact of bilingually instructed students upon a previously monolingual mid school or high school, and so on.

Obligatory comparisons.

The Title VII regulations also stipulate that the projects must report pre- and post-test results in reading for all project students and all comparison group students where such a group is used. This report must use "mean scores, standard deviations, and appropriate tests of statistical significance," Without a comparison "group," a project can compare its students with the populations on which the reading tests were normed. When a comparison group is identified, it can consist of earlier students at the same school, students in nonproject classrooms at the same or other locations, or some combination of these.

While these regulations lock projects into a pre/post design using test data, they are not as rigorous as the required choice in Title I ESEA, for example, among norm referenced, criterion referenced, analysis of co-variance, and linear regression models of evaluation. Nevertheless, the Title VII regulations confront project directors with such decisions as what tests

to use, whether or not to establish a comparison group, the adequacy of any comparison group, the merits of using national norms for assessing reading achievement, the accessibility of data from years prior to the project and others. Local projects' responses to these matters vary. Often, especially in smaller districts, they seek expert help from outside consultants who interpret the design requirements and frequently devise, help carry out, and report the evaluation. The more rigorous the technical specifications, the more emphatically the project director is tempted to alienate himself or herself from the evaluation in favor of the professionals. Such an abdication simply encourages an underlying tendency to perceive evaluation as a separate--even irrelevant--task from the job of directing the bilingual education project.

Quality control?

The regulations appear to exercise some degree of quality control over project evaluations, and indirectly over projects themselves. A similar earlier attempt at such control will be recalled by Title VII veterans as "program auditing." Every project had to engage an outside auditor who reviewed the project's evaluation procedures in much the same way that a fiscal auditor reviews a company's financial records. I was a program auditor. I once had the odd experience of meeting some real fiscal auditors from a federal department who were reviewing operations of a school while I was engaged in a Title VII program audit there. We chatted in the hall. They were incredulous at the restricted scope of the program audit, and could scarcely conceal their disgust at the toothlessness of program audit findings.

Vacuousness was not the only problem with program audits. Project directors had an understandably difficult time comprehending the difference between evaluation and audit; I always thought several program auditors had the same difficulty. Since the funding requirements were quite explicit about auditing and not very explicit about evaluating, some project directors were careful about fulfilling audit requirements and relatively careless about their evaluations. The reason for beating this dead auditing horse is that we are likely to get into the same game with the same players moved to new positions. Instead of obscure and philosophically dubious auditing requirements, we will have obscure and philosophically tenuous evaluation requirements. Instead of a select group who really understands the audit process and fulfills the role for proposal and report writing purposes, we will have a select group who really understands the evaluation models and satisfies the technical requirements at proposal writing time. Project directors will not necessarily be any more comfortable with evaluation.

Inadequate evaluation.

A frequent inadequacy with the evaluations of bilingual education projects is their lack of memory. An especially vivid example was a Texas district that was funded for five years and produced a series of one-year evaluation reports that could not be cross-compared. For example, project results were reported as class and grade means one year, quartiles the next, and as average grade equivalency in another year. This one-year-at-a-time coverage was underscored by a high attrition rate in both project and comparison classes; there were relatively few nonproject

classes, and there was a large amount of variability in the classes in both groups. (In all other respects it was a strong evaluation!)

Each separate year's evaluation report was conscientiously done to meet the accountability demands of the funding agency, which in turn made its recycling decisions to fund the project again and again. If the results showed anything besides the obstacles to conducting evaluation in that district, they showed that "same rate of progress" was a minimally illuminating notion; both project and nonproject youngsters fared equally badly with respect to national norms. With each successive year the average grade equivalencies of the "senior" classes fell further behind their actual grade at testing time.

This particular project exemplified some of the consequences of concentrating attention upon reporting of retrospective data for accountability purposes, segregating evaluation from the management of the project, and annually but not cumulatively compiling data on a restricted range of effects. There were other problems not peculiar to evaluation. For example, the building principals were contributing to high attrition by promoting project pupils into nonproject classes. This problem continued year after year with ineffectual attempts by the project to remedy it, and no demand from the funding source that it be remedied. In the context of a neglectful if not hostile district, even an evaluation which studied wider effects and used data for its own management decisions might not have been able to contribute toward a successful project.

Some persuasive results.

There have been some promising reports of outcomes from bilingual education projects. Two rather well known ones I would like to talk about here are the Navajo-English project at Rock Point Community School in Arizona and the French-English one at St. Lambert School in Montreal. The results from these projects are persuasive that some instances of bilingual education are conducive not only to no deficit but to superior achievement on the part of those who participate. I would like to examine what the reports from these projects do and do not have in common with the task of evaluation in the context of Title VII.

Rock Point was itself a Title VII project that followed upon earlier bilingual efforts under Title I and still earlier efforts in the area of English as a Second Language. In its first few years under Title VII the project evaluation concentrated upon finding the percent of students in the various grades who met objectives on several criterion-referenced tests. In 1976 and again in 1978, results of numerous comparisons of project and nonproject students were reported by Rosier and Farella and by Rosier and Holm, respectively. The comparisons used sub-tests of the Stanford Achievement Test, the Metropolitan Achievement Test, and the Test of Proficiency in English as a Second Language.

In English reading, aspects of English proficiency, and in mathematics, project youngsters were roughly comparable with similar youngsters in other schools in the same Bureau of Indian Affairs Agency at the outset; all were quite far behind national norms for children of their ages. With succeeding years in the project, youngsters whose education (in literacy, in math) began in Navajo not only outscored their predecessors at

Rock Point and their counterparts at the other Agency schools who had not been so educated, they pulled even with at grade 5 and then passed at grade 6 the national norm in English reading for their age and grade.

What is persuasive about these results is that in repeated and varied comparisons over a long time span and in more than one subject area, the same pattern of results emerged. However, all these results were not the subject of the Title VII project's own evaluation reports, which mainly summarized the results of criterion referenced testing at intervals throughout the year. The project devoted most of its evaluation resources to monitoring progress toward curricular objectives in real time. Decisions that could be served by that monitoring included whether or not to review or beef up instruction in particular classrooms prior to the next testing.

To mount the study of Rock Point required access to data on the same standardized test from five Bureau of Indian Affairs schools with populations similar to Rock Point's, covering a time period from roughly 1972 to 1977. Even the 1972 data includes pre-project fourth graders, so that the scope of the study goes another four years or so back to, say, 1968, when those students were beginning school. The five control schools all had strong programs in teaching English as a Second Language, so that the comparison was finely focused on bilingual education as the independent variable. The opportunity for such coordinated testing in similar sites over a long period was provided by the Agency-wide policies of the Bureau of Indian Affairs, under whose administration the Rock Point school had been operated prior to becoming a contract school. While the Bureau gets no credit for the project, its at times maddeningly monolithic

operation proved convenient for evaluation purposes. Correspondingly, Title VII did not mount the study, although it supported the annual evaluations of the project, and of course funded the Rock Point project itself.

The St. Lambert project was described in Lambert and Tucker's (1972) book and numerous other articles. Briefly, one experimental group followed another through a French immersion program. Comparison groups from two Montreal school systems included monolingual English speakers on the one hand and monolingual French speakers on the other. The pilot and follow-up experimental groups were composed of initially monolingual speakers of English. These groups maintained a healthy parity in English with the English speaking comparison groups. (In terms of North American norms, they sometimes reached the 80th percentile in language arts, for example.) At the same time, they performed quite creditably with respect to the French comparison groups. Although their spoken French, for example, was never indistinguishable from that of natives, it was proficient, and it was better than that of at least some natives. In mathematics, the experimental groups scored as well as the comparison groups even though they were taught through their second language (French) and tested in both languages. On measures of attitude the students in the experimental groups appeared ready to identify with both English and French cultures and eager to pursue their bilingual schooling.

The persuasiveness of the results from St. Lambert rests mainly on the replication built into the experiment; findings from the pilot group were mirrored quite closely in the follow-up group. The great number and variety of tests also contributed, but the availability of city-wide

achievement test norms from the Montreal (French) Catholic school system was especially helpful to the credibility of the results. To mount the St. Lambert study required considerable measurement expertise and linguistic knowledge, for example, in devising and scoring fluency tests, decoding tests. It also required access to comparison populations in similar neighborhoods, and to tests and test results, over a several year period.

The two school systems in Montreal allowed the same kind of coordination for evaluating St. Lambert that the Chinle Agency made possible in the case of Rock Point. None of these agencies actually managed the studies, (much less the projects); that role was played by university or quasi-university staff from McGill in Montreal, and by project and school staff (or former staff turned student) in the case of Rock Point. While these studies were clearly evaluative of the projects in question, in Stufflebeam's (1974) terms they represented research efforts yielding generalizable knowledge rather than project evaluations.[5] Their boundaries extended well beyond the scope of the management of the respective projects.

Projects outweigh evaluations.

Let us not lose sight of the project in discussing these studies and comparing them with project evaluation. It seems plausible that successful projects stand the best chance of yielding persuasive data. No one can state with certainty what conditions contribute to a local project's success; the list of such conditions could be long. Perhaps they include a strong desire in a community to change the use of two languages; local control; a population of students that is

homogeneous with respect to language dominance and proficiency; educational leadership with experience in bilingual education and ancillary fields; and continuity of effort. When these conditions (or some others that turn out to be more accurate) are met, the project should have a good chance for success. In that case the project's evaluation resources should be directed at assuring that success, not at grinding out a success story.

Characteristics of an evaluation aimed at project success.

A project evaluation aimed at securing the project's success would, for example, monitor the conditions just described, partly to establish their existence in the first place, and partly to help the project managers to keep those conditions favorable. Some kinds of data to collect might include, who wants to enroll in the bilingual project or leave it; who wishes to transfer in; who would send a younger brother or sister of an enrolled student; who wishes such a program for older brothers or sisters, and so on. An evaluation designed to serve the project's chances for success would assess the qualifications and inclinations of teachers to teach bilingually before extending a project vertically or horizontally. This kind of evaluation would establish a system for following the same students within the project over a span longer than one or two years. Such an evaluation would provide the capability of monitoring students' performance at frequent intervals during the school year, and of promptly reporting the results to staff members who can make use of them. Such an evaluation would assemble and interpret applicable data that is collected by others, for exam-

ple, results of district-wide or state-wide reading tests for the schools of interest, attendance records. An evaluation intended to improve the outlook for a successful project would also gauge cumulative (end of year) progress by students in the bilingual classes, with special attention to outcomes that the schools would not ordinarily be testing anyway, such as, say, writing proficiency in the other language, attitudes toward English.

All these kinds of evaluation activity merge into management; project directors investigate these matters anyhow, not necessarily under the heading of evaluation. Directors do these things not to account to the funding agency or to persuade the public, although these functions may be served incidentally; they investigate these matters so as to have a better project. The Title VII resources that go into local project evaluation should be used to amplify the capabilities of project directors to find out and use what they need to know in order to have a better project.

Reconstituting general evaluations.

To set up the kinds of studies that will inform national policymaking--and also impinge on public opinion on a larger scale than locally--requires coordination that extends beyond the local project in time and space. Rather than tightening the screws on local evaluations, one way to arrive at such informative studies would be to encourage attempts at coordination. We could encourage regional combinations that imitate just the facilitative properties of, say, the Chinle Agency in the case of Rock Point or the Montreal Catholic Schools in the case of St. Lambert. Some large school systems may already have the

kinds of resources that a superordinate organization lends to evaluative efforts. Even within a large district however, a very few schools may be involved in bilingual education. Those schools may be more relevantly compared to certain schools in other districts than to the remainder of their home district.

With the present regulations, each project director seeks his own method of meeting the requirements of reporting comparative performance. The Office of Education should recognize that some of the projects are more favorably situated for well designed evaluations than others. For some districts, the prospect of obtaining persuasive evaluation data for the centralized policy aspects of bilingual education could simply be dropped. A few (say 6 to 10) projects or clusters of projects could be selected where conditions for such informative evaluation studies were favorable. Title VII money could be used to support those several studies, to facilitate the coordination involved, for example. This latter possibility appeals to me much more than does letting one huge contract for one ostensibly national evaluation which then becomes vulnerable to all sorts of temptation and threats to validity.[6]

Advice.

My advice to directors of local projects in the meantime would be:

(1) Align your project with a collection of several schools, roughly as the members of the Northern New Mexico Consortium for Rural Education have done. Attempt to coordinate your selection of tests, times of testing, and methods

of analysis and reporting, much as the Chinle Agency incidentally did for Rock Point.

(2) Select the comparison group or groups with great care; specifically avoid a comparison group that is multiply advantaged with respect to the project participants. Some districts place all children who presumably could benefit from bilingual education into the project and no comparable children remain. In that case, either use no separate group, or seek a suitable group outside the district, among the other schools or districts with which the project is aligned, assuming item (1) above is accomplished.

(3) Keep the selection of tests, times, methods and so on the same for as many years as possible, even if the tests have flaws from the standpoint of the language and culture of students in your project. If you must change tests, give both the old and the new to at least one cohort of students so there is a basis for comparing results before and after the change.

(4) Distinguish evaluation activities that are to fulfill requirements of accountability from those that inform project managers. Then as far as possible, utilize the data collected for the former purpose to serve the latter. For example, all project and all comparison group students must be tested pre- and post- in reading. Decide which subdivisions of the project are most meaningful. Then when setting up the required testing, arrange to get test and subtest scores by the subdivisions that will reveal the most helpful information. These divisions are not necessarily the ones that a test scoring service ordinarily provides. Examples are children in different locations within a project, children who have been with the project all along or who entered late or whose

participation was interrupted, children who were highly proficient in the other language at the outset or were not, children who were team-taught bilingually by two monolinguals or were taught by one bilingual teacher. Some similar divisions may exist within nonparticipating comparison groups as well. Some possibilities are to compare project subgroups with each other or with the overall group, or with subgroups of a comparison group.

(5) Exploit as much as possible any tests that are already given under other auspices to pupils in your district. For example, New Mexico gives the Comprehensive Test of Basic Skills (CTBS) state-wide in grades 5, 8, and 11. Even for bilingual education projects that enroll only grades 1 through 4, this state-wide testing can provide a useful baseline. In two years, project students who are now in grades 3 and 4 will have grade 5 CTBS scores that can be compared with those of pre-project fifth graders, for example.

(6) Cool down everyone's expectations. It is usually not plausible that two years after they enter a bilingual project, the participants will display dramatic gains over nonparticipants; promising such gains puts the director on the defensive, and makes audiences skeptical.

Notes

1. See Epstein, pp. 9-31

2. U. S. Commissioner of Education, pp. 95-103

3. Stufflebeam, pp. 121, 122

4. Mackey, pp. 229-231

5. Stufflebeam, p. 120

6. See O'Malley

References

"Bilingual Education Program," Federal Register, 41:114, Friday, June 11, 1976, pp. 23860-23872.

Patricia Lee Engle, "Language medium in early school years for minority language groups," Review of Educational Research, 45:2, Spring, 1975, pp. 283-325.

Noel Epstein, Language, Ethnicity, and the Schools, Institute for Educational Leadership, George Washington University, Washington, D.C., 1977.

Executive Summary of the Final Report: Evaluation of the Impact of ESEA Title VII Bilingual Education Programs (The AIR Report), U. S. Department of Health, Education, and Welfare, Office of Education, Office of Planning, Budgeting and Evaluation, Washington, D. C., March, 1978.

Wallace E. Lambert and G. Richard Tucker, Bilingual Education of Children: The St. Lambert Experiment, Newbury House Publishers, Inc., Rowley, Massachusetts, 1972.

James B. Macdonald, "An evaluation of evaluation," The Urban Review, 7:1, January, 1974, pp. 3-14.

William F. Mackey, "The evaluation of bilingual education," in Bernard Spolsky and Robert L. Cooper, eds., Frontiers of Bilingual Education, Newbury House Publishers, Inc., Rowley, Massachusetts, 1977, pp. 226-281.

J. Michael O'Malley, "Review of the evaluation of the impact of ESEA Title VII Spanish/English bilingual education program," Bilingual Resources, 1:2, Winter, 1978, pp. 6-10.

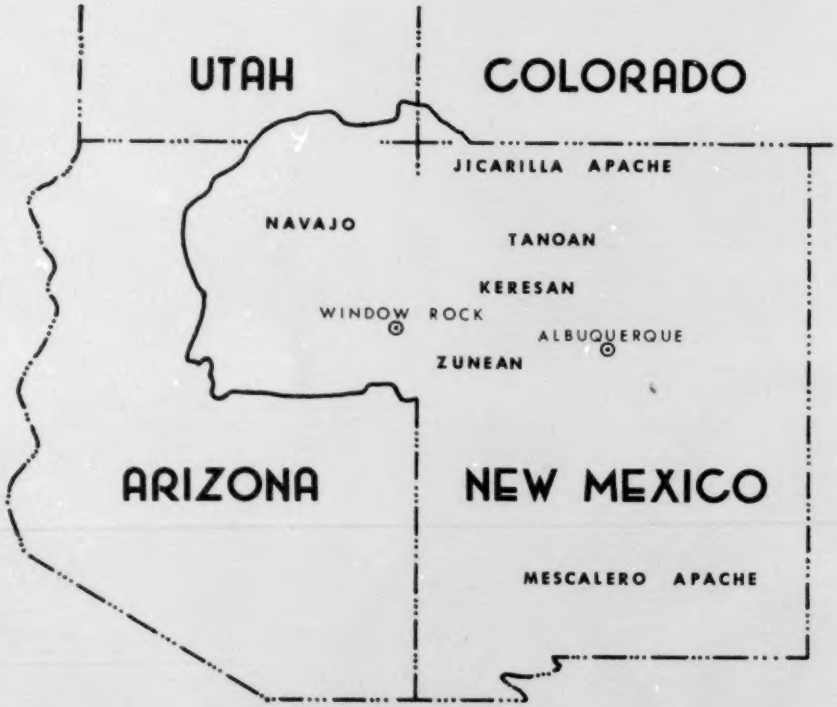
Paul Rosier and Merilyn Farella, "Bilingual education at Rock Point--Some early results," TESOL Quarterly, 10:4, December, 1976, pp. 379-388.

Paul Rosier and Wayne Holm, "Bilingual education for Navajo children: The Rock Point experience," mimeo, 1978.

Daniel L. Stufflebeam, "Alternative approaches to educational evaluation: A self study guide for educators," in W. James Popham, ed., Evaluation in Education: Current Applications, McCutchan Publishing Corporation, Berkeley, California, 1974, pp. 116-143.

U. S. Commissioner of Education, The Condition of Bilingual Education in the Nation: First Report by the U. S. Commissioner of Education to the President and the Congress, U. S. Department of Health, Education, and Welfare, Washington, D. C., 1977.

THE AIBEC SERVICE AREA



The American Indian Bilingual Education Center is part of the National Network of Centers for Bilingual Education. The Center provides support services in Bilingual/Bicultural Education to American Indian Local Education Agencies (LEA's), Institutions of Higher Education (IHE's), State Educational Agencies (SEA's) and other educational programs. These support services include programmatic and technical assistance as well as teacher training for E.S.E.A. American Indian Title VII projects. In addition to providing such services, the Center serves as a link between the LEA's, MDC's, and DAC's.